

Open Research Online

The Open University's repository of research publications and other research outputs

Dynamics Of Holomorphic Functions In The Hyperbolic Plane

Thesis

How to cite:

Christodoulou, Argyrios (2020). Dynamics Of Holomorphic Functions In The Hyperbolic Plane. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 2019 Argyrios Christodoulou



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Version of Record

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.21954/ou.ro.000111bb>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

Dynamics of holomorphic functions in the hyperbolic plane

Argyrios Christodoulou

BSc (Aristotle University of Thessaloniki)

MSc (Aristotle University of Thessaloniki)

A thesis submitted to

SCHOOL OF MATHEMATICS AND STATISTICS

THE OPEN UNIVERSITY

for the degree of Doctor of Philosophy

April 7, 2020

Declaration

I confirm that the material contained in this thesis is the result of independent work, except where explicitly stated, with the exception of Chapters 2 & 3 which are the result of joint work with Ian Short, and Chapters 5 & 6 which is the result of joint work with Ian Short and Matthew Jacques. None of it has previously been submitted for a degree or other qualification to this or any other university or institution.

Argyrios Christodoulou

April 8, 2020

Abstract

This thesis investigates the interactions between hyperbolic geometry and the dynamical behaviour of compositions of holomorphic self-maps of the hyperbolic plane. Our analysis draws inspiration from iteration theory and the theory of discrete groups.

First, we prove an inequality that quantifies how close a holomorphic function is to being a conformal self-map of the hyperbolic plane. This can be thought of as a rigidity result for conformal functions that involves the hyperbolic metric.

We then use our rigidity result in order to study the dynamics of holomorphic self-maps of the hyperbolic plane. In particular, we investigate the behaviour of compositions of a sequence of functions that itself converges to some limit function. Our goal is to examine under what conditions the dynamics of the composition sequence is similar to the dynamics of the iterates of the limit function. Intuitively, this question is about whether the Denjoy–Wolff theorem is stable under perturbations in the space of holomorphic functions.

Next, we focus on compositions of Möbius transformations. Due to a result of Jacques and Short, the dynamical behaviour of any composition sequence generated by finitely many Möbius transformations can be inferred from the topological properties, in the Möbius group, of the semigroup that these transformations generate. We introduce geometric conditions on these semigroups that allow us to interpret their topological behaviour.

Finally, we use our analysis of semigroups of Möbius transformations in order to study the parameter space of uniformly hyperbolic $\mathrm{PSL}(2, \mathbb{R})$ -cocycles. This topic was previously investigated by Avila, Bochi and Yoccoz, who proved that the uniform hyperbolicity of cocycles is equivalent to certain geometric properties of Möbius transformations in $\mathrm{PSL}(2, \mathbb{R})$. The three authors pose several questions about the structure of this parameter space and we provide answers to two of their questions.

Publications

Much of this thesis has been previously published in the form of papers, as follows:

- (1) The results in Chapter 2 have appeared in *Annales Academiæ Scientiarum Fennicæ Mathematica* [17]
- (2) The results in Chapter 3 have been submitted for publication (arxiv: 1907.09366)
- (3) The results in Chapter 4 have been submitted for publication (arxiv: 1907.00347)

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr Ian Short, for his continued guidance and support. I would also like to thank my master's thesis supervisor, Prof Dimitios Betsakos, for encouraging me to start a PhD and for all his advice throughout the years.

From the School of Mathematics and Statistics, I would like to thank Prof Phil Rippon and Prof Gwyneth Stallard for being a constant source of advice, Dr Matthew Jacques for all the interesting discussions, and Dr Vasiliki Evdoridou for her tremendous help.

I am very thankful to Evgenia for her immense support and encouragement, and her patience throughout this ordeal. Without her, the pressure would have been unbearable. Last but not least, I would like to thank my family for allowing me to pursue the things I am passionate for and for supporting my every decision.

Contents

Declaration

Abstract

Publications

Acknowledgements

Chapter 1. Introduction	1
1.1. Hyperbolic geometry and the Schwarz–Pick lemma	1
1.2. Iteration theory	6
Chapter 2. A hyperbolic-distance inequality for holomorphic maps	9
2.1. Statement of the main result	9
2.2. Holomorphic maps with a fixed point	14
2.3. Holomorphic maps of the disc	17
2.4. Holomorphic maps of the punctured disc	18
2.5. Theorem 2.1 in geometric function theory	20
Chapter 3. Stability of the Denjoy–Wolff theorem	23
3.1. Introduction and background	23
3.2. Stability at elliptic transformations and the identity function	25
3.3. Denjoy–Wolff point inside the disc	29
3.4. Denjoy–Wolff point on the boundary of the disc	31
Chapter 4. Universal constraints on semigroups of hyperbolic isometries	39
4.1. Semigroups of hyperbolic isometries	39
4.2. Examples and preliminary results	44
4.3. Two-generator semigroups	49

CONTENTS

4.4. Proof of Theorem 4.6	60
Chapter 5. Möbius semigroups and uniform hyperbolicity	65
5.1. The hyperbolic locus	65
5.2. The semidiscrete and inverse-free locus	70
5.3. The complement of \mathcal{S}	71
Chapter 6. Boundaries of the hyperbolic components	85
6.1. Outline of the chapter	85
6.2. Schottky semigroups	85
6.3. Semigroups within groups	88
Chapter 7. Open questions and future work	97
7.1. Right-compositions of holomorphic functions	97
7.2. Uniform hyperbolicity	97
Bibliography	

CHAPTER 1

Introduction

This thesis studies the dynamical properties of holomorphic functions that map the hyperbolic plane into itself. In particular, we investigate the geometric properties of holomorphic functions in the hyperbolic plane in order to extend classical results from the iteration theory of holomorphic functions and dynamical systems. Our analysis incorporates ideas from hyperbolic geometry, function theory and the theory of discrete groups. In this chapter we introduce the basic material from these areas of mathematics that will be used throughout the thesis. The more advanced techniques will be introduced as necessary in the following chapters.

1.1. Hyperbolic geometry and the Schwarz–Pick lemma

We denote the complex plane by \mathbb{C} and the extended complex plane $\mathbb{C} \cup \{\infty\}$ by $\overline{\mathbb{C}}$. Throughout the thesis, the identity function $z \mapsto z$ will be denoted by Id . We will be mostly use two models for the hyperbolic plane, which we now describe. Let $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ be the unit disc in the complex plane. The distance $\rho_{\mathbb{D}}$ induced by the Riemannian metric $\lambda_{\mathbb{D}}(z)|dz|$, where

$$\lambda_{\mathbb{D}}(z) = \frac{2}{1 - |z|^2},$$

is called the *hyperbolic distance in \mathbb{D}* . The metric space $(\mathbb{D}, \rho_{\mathbb{D}})$ is complete and is called the *disc model* of the hyperbolic plane. The geodesics in the disc model of the hyperbolic plane are circular arcs in \mathbb{D} that are tangent to the unit circle, and Euclidean lines through the origin.

The *upper half-plane model* of the hyperbolic plane is defined as follows. We denote by $\mathbb{H} = \{z \in \mathbb{C} : \text{Im } z > 0\}$ the upper half-plane in \mathbb{C} . The hyperbolic metric $\rho_{\mathbb{H}}$ in \mathbb{H} is

induced by the Riemannian metric $\lambda_{\mathbb{H}}(z)|dz|$, where

$$\lambda_{\mathbb{H}}(z) = \frac{1}{\operatorname{Im} z},$$

and the geodesics in $(\mathbb{H}, \rho_{\mathbb{H}})$ are either vertical half-lines or circular arcs perpendicular to the real line \mathbb{R} .

Using the Riemann mapping theorem, we can transfer the hyperbolic metric from \mathbb{D} to any simply connected, proper subdomain of \mathbb{C} . Let $D \subset \mathbb{C}$ be a simply connected domain. The hyperbolic distance ρ_D in D is induced by the Riemannian metric $\lambda_D(z)|dz|$, where

$$\lambda_D(z) = \lambda_{\mathbb{D}}(g(z))|g'(z)|,$$

where g is a Riemann map from D to \mathbb{D} . It is easy to check that λ_D does not depend on the choice of the function g (see [10, page 25]). To simplify notation we will often omit the subscript D from the distance ρ_D when there is no chance of confusion.

One of the most useful results concerning the hyperbolic metric in a simply connected domain D is that holomorphic self-maps of D contract hyperbolic distances between points. This is the famous Schwarz–Pick lemma (see, for example, [1, Theorem 1.1.6], [10, Theorem 6.4] or [16]).

THEOREM 1.1 (Schwarz–Pick). *Let D be a simply connected, proper subdomain of \mathbb{C} , and let f be a holomorphic function that maps D into itself. For any $z, w \in D$ we have*

$$\rho(f(z), f(w)) \leq \rho(z, w),$$

with equality for a pair $z, w \in D$ if and only if the function f is a conformal automorphism of D .

By the term *conformal automorphism* of D we mean a one-to-one holomorphic function from D onto itself. So in the case where our domain is either \mathbb{D} or \mathbb{H} , the Schwarz–Pick lemma tells us that the set of orientation-preserving isometries of the hyperbolic plane is the group of Möbius transformations that map \mathbb{D} or \mathbb{H} , respectively, onto itself. In particular, the group of conformal automorphisms of \mathbb{H} is exactly the group of real Möbius transformations

$$\operatorname{PSL}(2, \mathbb{R}) = \left\{ z \mapsto \frac{az + b}{cz + d} : a, b, c, d \in \mathbb{R} \quad \text{and} \quad ad - bc = 1 \right\},$$

(see [28, Section 1.1] or [39, Section 10.2]). The group $\mathrm{PSL}(2, \mathbb{R})$ acts on the extended real line $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$, and its non-identity elements are classified as *elliptic*, *parabolic* or *hyperbolic* depending on whether they have one fixed point in \mathbb{H} , one fixed point in $\overline{\mathbb{R}}$, or two fixed points in $\overline{\mathbb{R}}$, respectively.

Suppose that f is a hyperbolic transformation in $\mathrm{PSL}(2, \mathbb{R})$. The fixed point $\alpha(f)$ of f with $|f'(\alpha(f))| < 1$ is called *the attracting fixed point* of f . The other fixed point of f is called the *repelling fixed point*, and will be denoted by $\beta(f)$. Note that since f is an isometry of the hyperbolic metric, it preserves the unique geodesic that joins $\alpha(f)$ and $\beta(f)$. This geodesic is called the *axis* of f and is denoted by $\mathrm{Ax}(f)$. In addition, the *translation length* of f is the distance $\rho(w, f(w))$, for any $w \in \mathrm{Ax}(f)$, and will be denoted by $\tau(f)$. The translation length of f can be thought of as a measure of how much the transformation distorts points in the hyperbolic plane.

If f and g are transformations in $\mathrm{PSL}(2, \mathbb{R})$, then the *commutator* of f and g is defined to be the transformation $[f, g] = f \circ g \circ f^{-1} \circ g^{-1}$. Also, for a transformation $h(z) = (az + b)/(cz + d)$ with $ad - bc = 1$, we define the *trace* of h to be the number $\mathrm{tr}(h) = a + d$. The trace of a transformation provides us with the following classification of elements in $\mathrm{PSL}(2, \mathbb{R})$.

Let h be a transformation in $\mathrm{PSL}(2, \mathbb{R})$. Then h is:

- (1) parabolic or the identity if and only if $|\mathrm{tr}(h)| = 2$;
- (2) elliptic if and only if $|\mathrm{tr}(h)| < 2$; and
- (3) hyperbolic if and only if $|\mathrm{tr}(h)| > 2$.

In our figures, a hyperbolic transformation will be portrayed by drawing its axis as a directed hyperbolic line pointing towards its attracting fixed point (see Figure 1.1 on the left). A parabolic transformation will be portrayed as a directed Euclidean circle, of arbitrary Euclidean radius, that is tangent to \mathbb{R} at the fixed point of the parabolic transformation. The direction of the circle indicates the action of the transformation on \mathbb{H} (see Figure 1.1 on the right).

An immediate consequence of the definition of the hyperbolic metric is that it is conformally invariant. That is, if D_1 and D_2 are simply connected domains and ϕ is a conformal map from D_1 onto D_2 , then $\rho_{D_2}(\phi(z), \phi(w)) = \rho_{D_1}(z, w)$, for all $z, w \in D_1$.

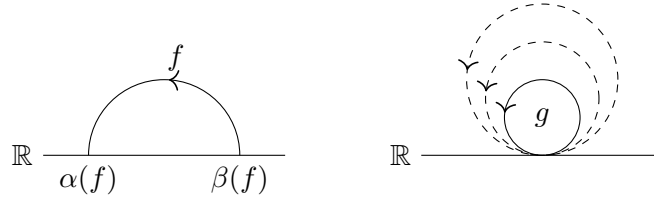


FIGURE 1.1. A hyperbolic transformation f and a parabolic transformation g .

[10, Theorem 6.3]. This allows us to move freely between different models of the hyperbolic plane. For the disc and upper half-plane models, it is handy to use the Cayley transform

$$C(z) = \frac{z - i}{z + i},$$

which is a conformal map from \mathbb{H} onto \mathbb{D} , and thus a hyperbolic isometry. If h is a conformal isometry of the hyperbolic metric in \mathbb{D} , then h will be called *elliptic*, *parabolic* or *hyperbolic*, if the transformation $C^{-1} \circ h \circ C \in \text{PSL}(2, \mathbb{R})$ is elliptic, parabolic or hyperbolic, respectively.

If Γ is a smooth curve in a simply connected, proper subdomain D of \mathbb{C} , then for any point $z \in D$, we let $\rho_D(z, \Gamma)$ be the hyperbolic distance of z from the curve Γ , in D . That is

$$\rho_D(z, \Gamma) = \inf\{\rho_D(z, w) : w \in \Gamma\}.$$

We now transfer the hyperbolic metric to domains that are not simply connected. A domain S in \mathbb{C} will be called a *hyperbolic domain* if its complement in \mathbb{C} contains at least two points. Obviously, any simply connected, proper subdomain of \mathbb{C} is a hyperbolic domain.

For any domains A and B in \mathbb{C} , a function $h: A \rightarrow B$ will be called a *holomorphic covering*, if for every point $z \in B$ there exists an open neighbourhood U of z , such that $f^{-1}(U) = \bigcup V_i$ is a disjoint union of open sets V_i , such that the restriction of f to each V_i is a conformal map from V_i onto U . So, holomorphic coverings can be thought of as holomorphic functions that are locally conformal. Also, trivially, all conformal maps are holomorphic coverings.

For hyperbolic domains we have an analogue of the Riemann mapping theorem, called the

uniformization theorem, which we now state (see, for example, [10, Theorem 10.2] or [37, Theorem 2.1]).

THEOREM 1.2 (Uniformization theorem). *A domain $S \subset \mathbb{C}$ is a hyperbolic domain if and only if there exists a holomorphic covering $h: \mathbb{D} \rightarrow S$.*

The function h in Theorem 1.2 is called a universal covering of S . The existence of a holomorphic covering in the uniformization theorem allows us to transfer the hyperbolic metric from \mathbb{D} to any hyperbolic domain S in the following way. Let λ_S be the pull-back of $\lambda_{\mathbb{D}}$ under a universal covering of S . Then λ_S is a complete Riemann metric of constant curvature -1 [10, Theorem 10.3], and in fact it is the unique metric on S with these properties. We will call λ_S the *hyperbolic metric in S* .

The inequality of the Schwarz–Pick lemma still holds for any hyperbolic domain, that is every holomorphic self-map of a hyperbolic domain contracts hyperbolic distances, but determining those holomorphic functions, and those pairs of points, for which we have equality is more complicated (see [10, Theorem 10.5]). Note that since the hyperbolic metric is defined in terms of the holomorphic covering h between two hyperbolic domains, h is a local isometry of the hyperbolic metric in these domains [10, Theorem 10.4].

The *fundamental group* of a hyperbolic domain S is defined as

$$\pi_1(S) = \{\gamma: \mathbb{D} \rightarrow \mathbb{D}: \gamma \text{ is a conformal automorphism of } \mathbb{D} \text{ and } h \circ \gamma = h\},$$

where h is a universal covering of S . The definition of the fundamental group of a hyperbolic domain that we presented is equivalent to the usual definition in algebraic topology (i.e. the group of homotopy classes of closed curves in the domain) [21, Theorem 5.6]. Also, $\pi_1(S)$ acts transitively on the fibres of the universal covering of S ; that is, for any point $z \in S$ and any pair of points $w_1, w_2 \in h^{-1}(z)$, there exists $\gamma \in \pi_1(S)$ such that $\gamma(w_1) = w_2$. For more information on the fundamental groups of domains, we refer to [1, Section 1.1.3] and [21, Chapter 1].

In the second chapter of this thesis, we use the Schwarz–Pick lemma and elementary properties of the hyperbolic metric in order to prove a rigidity result for holomorphic self-maps of the hyperbolic plane.

1.2. Iteration theory

For a function f holomorphic on some subset of \mathbb{C} , we define the n^{th} *iterate* of f to be the function

$$f^n = \underbrace{f \circ f \circ \cdots \circ f}_{n \text{ times}}.$$

For a domain $D \subseteq \mathbb{C}$, we denote by $\text{Hol}(D)$ the set of holomorphic functions that map D into itself. The usual topology on the space of functions $\text{Hol}(D)$ is the compact-open topology. A sub-base in this topology consists of sets of the form

$$[K, U] = \{f \in \text{Hol}(D) : f(K) \subset U\},$$

where K is compact and U is open. To be more specific, a set in the compact-open topology of $\text{Hol}(D)$ is open if and only if it is a union of sets of the form

$$[K_1, U_1] \cap [K_2, U_2] \cap \cdots \cap [K_m, U_m],$$

where K_i is compact and U_i is open, for all $i = 1, 2, \dots, m$. We may also endow $\text{Hol}(D)$ with the topology of locally uniform convergence. We say that a sequence of functions (f_n) in $\text{Hol}(D)$ *converges locally uniformly* to a function f if for every compact subset K of D we have that

$$\sup_{z \in K} \chi(f_n(z), f(z)) \xrightarrow{n \rightarrow \infty} 0,$$

where χ is the chordal metric in the extended complex plane $\overline{\mathbb{C}}$. By Hurwitz's theorem, the limit function f can either be a constant in \overline{D} , or a non-constant function in $\text{Hol}(D)$. It is a classical result that the compact-open topology and the topology of locally uniform convergence coincide on any domain D (for more details see [31, p. 221]). So from now on, whenever we say that a sequence of functions is convergent, we mean that it is locally uniformly convergent. We will, however, use the open sets of the compact-open topology in order to state several of our results in Chapter 3.

In this thesis, we will mostly be interested in holomorphic self-maps of the hyperbolic plane, i.e. the space of functions $\text{Hol}(\mathbb{D})$ or $\text{Hol}(\mathbb{H})$. One of the most celebrated theorems in the iteration theory of holomorphic functions is the Denjoy–Wolff theorem, which provides us with a complete description for the dynamics of the iterates of a holomorphic function (see, for example, [37, Theorem 5.4]).

THEOREM 1.3 (Denjoy–Wolff). *Suppose that f is a holomorphic self-map of the unit disc \mathbb{D} . Then either*

- (1) *f is the identity function or an elliptic Möbius transformation that fixes \mathbb{D} , or*
- (2) *there exists a point $\zeta \in \overline{\mathbb{D}}$ such that the sequence of iterates (f^n) converges locally uniformly on \mathbb{D} to ζ .*

The point ζ will be called *the Denjoy–Wolff point of f* . If the Denjoy–Wolff point ζ of f lies in \mathbb{D} , then it is the unique fixed point of f . However, if ζ lies on the unit circle, then f has no fixed points in \mathbb{D} and $f(z) \rightarrow \zeta$, as z tends to ζ through an angular sector at ζ . For this reason, the Denjoy–Wolff point is often called the attracting boundary fixed point of f .

The behaviour of a function f around its Denjoy–Wolff point has been thoroughly investigated due to its many connections with other topics in complex analysis, such as composition operators [44], ergodic theory [14] and inner functions [20], among others. There have also been many attempts at generalising the Denjoy–Wolff theorem in order to allow for compositions of multiple functions, instead of just the iterates of one map. In Chapter 3 we will present an overview of these results, and we will examine whether the Denjoy–Wolff theorem is stable under perturbations of the function f in the space $\text{Hol}(\mathbb{D})$.

Observe that if f is a hyperbolic transformation in $\text{PSL}(2, \mathbb{R})$ then the n^{th} iterate of f is also a hyperbolic transformation, with translation length $n\tau(f)$. Furthermore, as we saw earlier, the set $\text{PSL}(2, \mathbb{R})$ lies in the space $\text{Hol}(\mathbb{H})$, and thus carries the topology of locally uniform convergence.

A subset of $\text{PSL}(2, \mathbb{R})$ will be called *discrete* if it is a discrete set in the topology of $\text{PSL}(2, \mathbb{R})$. An interesting example of a discrete subgroup of $\text{PSL}(2, \mathbb{R})$ is the fundamental group of a hyperbolic domain S , where the covering map is chosen as $h: \mathbb{H} \rightarrow S$ [1, Lemma 1.1.23]. The study of discrete subgroups of $\text{PSL}(2, \mathbb{R})$ is intertwined with several topics in hyperbolic geometry (see for example [6] and [28]) and in Chapters 4, 5 and 6 we explore how this connection between the dynamics of Möbius transformations and hyperbolic geometry extends to semigroups in $\text{PSL}(2, \mathbb{R})$.

Finally, suppose that $f \in \text{Hol}(D)$ is a conformal function, where D is a subset of \mathbb{C} . We say that a set $A \subseteq D$ is *forward invariant under f* if $f(A) \subseteq A$. Similarly, A is *backward*

invariant under f if $f^{-1}(A) \subseteq A$. Also, A is *mapped compactly inside itself* by f if $\overline{f(A)} \subset A$.

CHAPTER 2

A hyperbolic-distance inequality for holomorphic maps

2.1. Statement of the main result

In this chapter, we are going to prove an inequality that quantifies the distance of a function $f \in \text{Hol}(\mathbb{D})$ from the identity function using two points $a, b \in \mathbb{D}$ and the hyperbolic distance in the unit disc. This estimate will prove particularly useful to our analysis of the stability of the Denjoy–Wolff theorem, in Chapter 3 to follow. Also, in the last section of this chapter we are going to describe how the inequality can be applied to provide an alternative proof of a well-known result in geometric function theory.

Our principal objective is to establish the following theorem.

THEOREM 2.1. *Suppose that f is a holomorphic self-map of \mathbb{D} and $a, b, z \in \mathbb{D}$, with $a \neq b$. Then*

$$\rho(f(z), z) \leq K(\rho(f(a), a) + \rho(f(b), b)),$$

where

$$K = \frac{\exp(\rho(z, a) + \rho(a, b) + \rho(b, z))}{\rho(a, b)}.$$

A somewhat similar result was obtained for Möbius transformations acting on the extended complex plane, using the chordal metric, in [11].

A strength of Theorem 2.1 is that the constant K is independent of the function f . As a consequence, we can use the theorem to prove quantitative versions of existing results about holomorphic maps close to the identity function. For instance, it is known that if (f_n) is a sequence of holomorphic self-maps of \mathbb{D} such that $f_n(a) \rightarrow a$ and $f_n(b) \rightarrow b$, for two distinct points a and b , then (f_n) converges locally uniformly on \mathbb{D} to the identity function. This can be proven by a normal families argument (see, for example, [43, Theorem 2.4.2]), but such an argument does not give an explicit rate of convergence. In contrast, Theorem 2.1 provides a rate of convergence, which allows us to make stronger

statements; for example, the theorem tells us that if the sum $\sum \rho(f_n(z), z)$ converges for $z = a, b$, then in fact it converges for *any* $z \in \mathbb{D}$.

If we fix two distinct points a and b in \mathbb{D} , then Theorem 2.1 (and the inequality $\rho(z, b) \leq \rho(z, a) + \rho(a, b)$) show us that there is a constant k depending only on a and b for which

$$(2.1.1) \quad \rho(f(z), z) \leq k e^{2\rho(z, a)} (\rho(f(a), a) + \rho(f(b), b)),$$

for all $z \in \mathbb{D}$ and all holomorphic self-maps f of \mathbb{D} . We now describe an example to show that the expression $e^{2\rho(z, a)}$ in this inequality cannot be significantly reduced.

For this example, we switch from \mathbb{D} to the right half-plane model of the hyperbolic plane, denoted by \mathbb{K} , and consider holomorphic self-maps of \mathbb{K} . We continue to denote the hyperbolic metric by ρ , on \mathbb{K} as on \mathbb{D} , and we make use of the formula $\rho(u, v) = \log(v/u)$, for points u and v on the positive real axis, with $u < v$.

Let $a = 1$ and $b = 2$. Let $f_n(w) = w + 1/n^2$ and $z_n = 1/n$, for $n = 1, 2, \dots$. Then $e^{\rho(z_n, a)} = n$, and

$$\rho(f_n(z_n), z_n) \sim \frac{1}{n}, \quad \rho(f_n(a), a) \sim \frac{1}{n^2}$$

(where, for two positive sequences (x_n) and (y_n) , we write $x_n \sim y_n$ to mean that there is a positive constant λ such that $x_n/\lambda < y_n < \lambda x_n$, for $n = 1, 2, \dots$). So

$$\frac{\rho(f_n(z_n), z_n)}{\rho(f_n(a), a)} \sim e^{\rho(z_n, a)}.$$

That is, the quotient of the distortion of f_n at z_n by the distortion of f_n at a grows exponentially with the hyperbolic distance between z_n and a . This examples indicates that the expression $e^{2\rho(z, a)}$ in inequality (2.1.1) cannot be made any smaller than $e^{\rho(z, a)}$.

The proof of Theorem 2.1 uses the fact that any holomorphic self-map of \mathbb{D} contracts the hyperbolic metric on \mathbb{D} , as stated in the Schwarz–Pick Lemma. We observe, however, that the theorem fails for the class of contractions of \mathbb{D} with the hyperbolic metric, which is broader than the class of holomorphic self-maps of \mathbb{D} . To see this, consider the function $f(w) = \operatorname{Re} w$, which contracts the hyperbolic metric on \mathbb{D} . Given any two distinct real numbers a and b in \mathbb{D} , we have $\rho(f(a), a) = \rho(f(b), b) = 0$, but $\rho(f(z), z)$ is positive for nonreal points z , so the inequality in Theorem 2.1 fails.

To illuminate later work, we record the following minor generalisation of Theorem 2.1. Recall that the conformal automorphisms of \mathbb{D} are hyperbolic isometries, and using this property we can immediately deduce the result from Theorem 2.1.

COROLLARY 2.2. *Suppose that f is a holomorphic self-map of \mathbb{D} , h is a conformal automorphism of \mathbb{D} , and $a, b, z \in \mathbb{D}$, with $a \neq b$. Then*

$$\rho(f(z), h(z)) \leq K(\rho(f(a), h(a)) + \rho(f(b), h(b))),$$

where $K = \rho(a, b)^{-1} \exp(\rho(z, a) + \rho(a, b) + \rho(b, z))$.

Theorem 2.1 and Corollary 2.2 could of course be stated with any simply connected domain in place of \mathbb{D} . Consider, now, a general hyperbolic domain S , and suppose that $\pi_1(S)$ is nonabelian. [1, Theorem 1.2.19] states that if S has a nonabelian fundamental group, then the identity function is an isolated point in the space $\text{Hol}(S)$. Observe that, unlike in simply connected domains, there might exist functions in $\text{Hol}(S)$, other than the identity, that fix two distinct points $z, w \in S$. However, [1, Corollary 1.3.8] states that all such functions are finite order automorphisms of S . Hence, we cannot obtain an inequality similar to that of Theorem 2.1 for finite order automorphisms of S . Also, it is easy to see that since Id is isolated in $\text{Hol}(S)$, the set of all finite order automorphisms of S is a discrete subset of $\text{Hol}(S)$.

Assume now that a and b are distinct points in S . We claim that there exists $\varepsilon = \varepsilon(a, b) > 0$ such that for every $f \in \text{Hol}(S)$, which is not a finite order automorphism of S , either $\varepsilon \leq \rho(f(a), a)$ or $\varepsilon \leq \rho(f(b), b)$. Assume, that for every $\varepsilon > 0$ there exists a function $f \in \text{Hol}(S)$, not a finite order automorphism of S , such that $\rho(f(a), a) \leq \varepsilon$ and $\rho(f(b), b) \leq \varepsilon$. This implies that we can find a sequence (f_n) in $\text{Hol}(S)$, such that f_n is not a finite order automorphism of S , for any n , $f_n(a) \rightarrow a$ and $f_n(b) \rightarrow b$, as $n \rightarrow \infty$. Because (f_n) is a normal family, we can assume that f_n converges to a function $f \in \text{Hol}(S)$, and since a and b are distinct points, f has to be a nonconstant function. Also, from the convergence of $f_n(a)$ and $f_n(b)$, we have that f fixes a and b , and thus it is either the identity or a finite order automorphism of S . This is a contradiction due to our discussion in the previous paragraph, and thus our claim is proved.

Suppose that $f \in \text{Hol}(S)$ is not a finite order automorphism of S . We are going to prove

that an inequality similar to that of Theorem 2.1 holds for f . To that end, note that there exists $\varepsilon > 0$ such that either $\varepsilon \leq \rho(f(a), a)$ or $\varepsilon \leq \rho(f(b), b)$, and the constant ε depends only on a and b . Thus, assuming that $\varepsilon \leq \rho(f(a), a)$, we have

$$\begin{aligned} \rho(f(z), z) &\leq \rho(f(z), f(a)) + \rho(f(a), a) + \rho(a, z) \\ &\leq 2\rho(a, z) + \rho(f(a), a) \\ &\leq \left(\frac{2\rho(a, z)}{\varepsilon} + 1 \right) \rho(f(a), a) \\ &\leq \left(\frac{2\rho(a, z)}{\varepsilon} + 1 \right) (\rho(f(a), a) + \rho(f(b), b)). \end{aligned}$$

A similar inequality would hold if $\varepsilon \leq \rho(f(b), b)$.

We conclude that in the case of hyperbolic domains with nonabelian fundamental group, one can certainly obtain a result similar to Theorem 2.1 for a subset of $\text{Hol}(S)$, but it is of little consequence since it follows easily from the topological properties of $\text{Hol}(S)$. This indicates that Theorem 2.1 is only significant for families of functions that come arbitrarily close to the identity function.

If S is a hyperbolic domain with abelian fundamental group, and is not simply connected, then it must be doubly connected [1, Theorem 1.1.29]; any such domain is conformally equivalent either to an annulus $A_r = \{z : 1/r < |z| < r\}$ (where $r > 1$) or to the punctured disc $\mathbb{D}^* = \mathbb{D} \setminus \{0\}$ [1, Corollary 1.1.30]. The conformal automorphisms of A_r are rotations, and rotations composed with the map $z \mapsto 1/z$. The remaining holomorphic self-maps of A_r are all homotopic in the family of continuous self-maps of A_r to constant maps [10, Corollary 13.7], and the set of these maps does not contain the identity function in its closure. Thus there is no worthwhile analogue of Theorem 2.1 for annuli either, since any sequence of holomorphic self-maps of A_r that converges to the identity function must eventually consist of rotations, and their geometry is straightforward.

The punctured disc \mathbb{D}^* is different, however, because there are plenty of nontrivial holomorphic self-maps of \mathbb{D}^* in any neighbourhood of the identity function. To consider these maps, we use the universal covering map $\pi: \mathbb{H} \rightarrow \mathbb{D}^*$ given by $\pi(\zeta) = e^{2\pi i \zeta}$, where \mathbb{H} is the upper half-plane. We use \mathbb{H} for the universal covering space rather than \mathbb{D} because

doing so gives a simpler covering map (and \mathbb{H} is also marginally easier to work with than \mathbb{D}).

Any holomorphic self-map f of \mathbb{D}^* lifts to a holomorphic self-map \tilde{f} of \mathbb{H} with $\pi \circ \tilde{f} = f \circ \pi$ (see [1, page 16]). This condition implies that the map \tilde{f} satisfies $\tilde{f}(\zeta + 1) = \tilde{f}(\zeta) + m$, for all $\zeta \in \mathbb{H}$ and some nonnegative integer m . The integer m is called the *degree* of f , and it is denoted by $\deg(f)$. It can also be defined using the formula

$$\deg(f) = \frac{1}{2\pi i} \int_{\gamma} \frac{f'(z)}{f(z)} dz,$$

where $\gamma(t) = \frac{1}{2}e^{2\pi it}$ ($t \in [0, 1]$). Since \deg is a continuous function, the set of holomorphic self-maps of \mathbb{D}^* of degree m , for any nonnegative integer m , is a closed set. The identity function belongs to the set of maps of degree 1. For more on the degree, see [10, Section 13].

The origin is a removable singularity for any holomorphic self-map f of \mathbb{D}^* , and if $\deg(f) > 0$, then the origin is fixed by f . It is reasonable, therefore, to expect to obtain a one-point inequality for self-maps of \mathbb{D}^* of positive degree akin to the earlier two-point inequalities. The next theorem is of this type; it is similar to Corollary 2.2, but the conformal automorphism of \mathbb{D} is replaced by a holomorphic self-covering map of \mathbb{D}^* . Such a map has the form $h(z) = e^{i\theta}z^m$, where $\theta \in \mathbb{R}$ and $m \in \mathbb{N}$.

In this theorem, $\lambda^*(z)|dz| = -|dz|/(|z|\log|z|)$ is the Riemannian metric on \mathbb{D}^* that gives rise to the hyperbolic distance ρ^* on \mathbb{D}^* .

THEOREM 2.3. *Suppose that f is a holomorphic self-map of \mathbb{D}^* , h is a holomorphic self-covering map of \mathbb{D}^* , and $\deg(f) = \deg(h) > 0$. Suppose also that $a, z \in \mathbb{D}^*$. Then*

$$\rho^*(f(z), h(z)) \leq L^3 \rho^*(f(a), h(a)),$$

where $L = 8\lambda^*(a) \exp \rho^*(z, a)$.

When h is the identity function and $\deg(f) = 1$, we obtain a one-point inequality comparable with Theorem 2.1.

COROLLARY 2.4. *Suppose that f is a holomorphic self-map of \mathbb{D}^* with $\deg(f) = 1$ and $a, z \in \mathbb{D}^*$. Then*

$$\rho^*(f(z), z) \leq L^3 \rho^*(f(a), a),$$

where $L = 8\lambda^*(a) \exp \rho^*(z, a)$.

2.2. Holomorphic maps with a fixed point

In this section we prove the following theorem, which is a version of Theorem 2.1 for holomorphic self-maps of the disc with a fixed point.

THEOREM 2.5. *Suppose that $a, b, z \in \mathbb{D}$, with $a \neq b$, and f is a holomorphic self-map of \mathbb{D} that fixes b . Then*

$$\rho(f(z), z) \leq M \rho(f(a), a),$$

where

$$M = \frac{\exp(\rho(a, z) + \rho(z, b))}{4 \sinh \frac{1}{2} \rho(a, b)}.$$

It suffices to prove this theorem when $b = 0$, as can be seen by conjugating f by a conformal automorphism of \mathbb{D} that takes b to 0. So let us assume, henceforth, that $b = 0$. We can also assume that $z \neq 0$, because, for $b = 0$, the inequality clearly holds when $z = 0$.

We recall two formulas for the hyperbolic metric on \mathbb{D} (see [10, page 15]), which state that, for $u, v \in \mathbb{D}$,

$$(2.2.1) \quad \sinh \frac{1}{2} \rho(u, v) = \frac{|u - v|}{\sqrt{(1 - |u|^2)(1 - |v|^2)}}, \quad \cosh \frac{1}{2} \rho(u, v) = \frac{|1 - u\bar{v}|}{\sqrt{(1 - |u|^2)(1 - |v|^2)}}.$$

The equations in the next lemma are merely special cases of these formulas, with $v = 0$.

LEMMA 2.6. *If $u \in \mathbb{D}$, then*

$$(1) \quad \sinh \frac{1}{2} \rho(u, 0) = \frac{|u|}{\sqrt{1 - |u|^2}},$$

$$(2) \quad \cosh \frac{1}{2} \rho(u, 0) = \frac{1}{\sqrt{1 - |u|^2}}.$$

We can use the equations in this lemma to replace the square-root terms from the left-hand formula from (2.2.1) in two ways, to give two more formulas involving the hyperbolic metric, presented in the following lemma.

LEMMA 2.7. *If $u, v \in \mathbb{D}$, then*

$$(1) \quad |u - v| = \frac{\sinh \frac{1}{2} \rho(u, v)}{\cosh \frac{1}{2} \rho(u, 0) \cosh \frac{1}{2} \rho(v, 0)},$$

$$(2) \quad \frac{|u - v|}{|u|} = \frac{\sinh \frac{1}{2} \rho(u, v)}{\sinh \frac{1}{2} \rho(u, 0) \cosh \frac{1}{2} \rho(v, 0)}.$$

We will apply Lemmas 2.6 and 2.7 repeatedly, so it is handy to define

$$s(u, v) = \sinh \frac{1}{2}\rho(u, v) \quad \text{and} \quad c(u, v) = \cosh \frac{1}{2}\rho(u, v).$$

Let us proceed with the proof of Theorem 2.5. We can assume that f is not a conformal automorphism of \mathbb{D} fixing the origin (a Euclidean rotation) because such maps are limits of sequences of holomorphic maps that are not conformal automorphisms (in the topology of compact convergence), and the inequality is preserved on taking this type of limit.

We define

$$g(w) = \begin{cases} f(w)/w, & w \neq 0, \\ f'(0), & w = 0. \end{cases}$$

Since $|f(w)| < |w|$ for $w \neq 0$, by Schwarz's lemma, we see that g is also a holomorphic map from \mathbb{D} to itself.

Recall that $a, z \in \mathbb{D} \setminus \{0\}$. Then $|1 - g(z)| \leq |1 - g(a)| + |g(a) - g(z)|$. That is,

$$\frac{|z - f(z)|}{|z|} \leq \frac{|a - f(a)|}{|a|} + |g(a) - g(z)|.$$

Applying Lemma 2.7 to this inequality, we obtain

$$\frac{s(f(z), z)}{s(z, 0)c(f(z), 0)} \leq \frac{s(f(a), a)}{s(a, 0)c(f(a), 0)} + \frac{s(g(a), g(z))}{c(g(a), 0)c(g(z), 0)}.$$

Since $c(f(z), 0) \leq c(z, 0)$, by the Schwarz–Pick lemma, and $c(f(a), 0) \geq 1$, we can rearrange this inequality to give

$$(2.2.2) \quad s(f(z), z) \leq s(z, 0)c(z, 0) \left(\frac{s(f(a), a)}{s(a, 0)} + \frac{s(g(a), g(z))}{c(g(a), 0)c(g(z), 0)} \right).$$

Next, observe that $\rho(g(z), 0) \geq |\rho(g(a), 0) - \rho(g(a), g(z))|$, so, since \cosh is an even function,

$$c(g(z), 0) \geq c(g(a), 0)c(g(a), g(z)) - s(g(a), 0)s(g(a), g(z)).$$

Multiplying both sides by $c(g(a), 0)$ and then applying the equations $c(g(a), 0)^2 = 1/(1 - |g(a)|^2)$ and $s(g(a), 0)c(g(a), 0) = |g(a)|/(1 - |g(a)|^2)$ (from Lemma 2.6), we see that

$$\begin{aligned} c(g(z), 0)c(g(a), 0) &\geq c(g(a), 0)^2 c(g(a), g(z)) - s(g(a), 0)c(g(a), 0)s(g(a), g(z)) \\ &= \frac{c(g(a), g(z)) - |g(a)|s(g(a), g(z))}{1 - |g(a)|^2} \\ &\geq \frac{c(g(a), g(z)) - |g(a)|s(g(a), g(z))}{1 + |g(a)|} \frac{|a|}{|a - f(a)|} \\ &= \frac{c(g(a), g(z)) - |g(a)|s(g(a), g(z))}{1 + |g(a)|} \frac{s(a, 0)c(f(a), 0)}{s(f(a), a)}, \end{aligned}$$

where, in the last line, we have applied Lemma 2.7 again. Rearranging this, we find that

$$\frac{1}{c(g(z), 0)c(g(a), 0)} \leq \frac{1 + |g(a)|}{c(g(a), g(z)) - |g(a)|s(g(a), g(z))} \times \frac{s(f(a), a)}{s(a, 0)}.$$

We now combine this inequality with (2.2.2) to give

$$(2.2.3) \quad s(f(z), z) \leq \frac{s(z, 0)c(z, 0)}{s(a, 0)} \left(1 + \frac{(1 + |g(a)|)s(g(a), g(z))}{c(g(a), g(z)) - |g(a)|s(g(a), g(z))} \right) s(f(a), a).$$

The part in large brackets is equal to

$$\frac{c(g(a), g(z)) + s(g(a), g(z))}{c(g(a), g(z)) - |g(a)|s(g(a), g(z))} \leq \frac{c(g(a), g(z)) + s(g(a), g(z))}{c(g(a), g(z)) - s(g(a), g(z))} \leq e^{\rho(a, z)},$$

where, for the last inequality, we applied the Schwarz–Pick lemma to g . Since $s(z, 0)c(z, 0) = \frac{1}{2} \sinh \rho(z, 0) \leq \frac{1}{4} e^{\rho(z, 0)}$, we see that (2.2.3) reduces to

$$(2.2.4) \quad s(f(z), z) \leq \frac{e^{\rho(a, z) + \rho(z, 0)}}{4s(a, 0)} s(f(a), a).$$

To finish, observe that Theorem 2.5 is clearly true if $\rho(f(z), z) < \rho(f(a), a)$, because $M > 1$. Assume then that $\rho(f(z), z) \geq \rho(f(a), a)$. The function $x \mapsto \sinh x/x$ is increasing for $x > 0$, as one can prove by differentiating it, so

$$\frac{\sinh \frac{1}{2} \rho(f(z), z)}{\frac{1}{2} \rho(f(z), z)} \geq \frac{\sinh \frac{1}{2} \rho(f(a), a)}{\frac{1}{2} \rho(f(a), a)}.$$

Hence

$$\frac{\rho(f(z), z)}{\rho(f(a), a)} \leq \frac{s(f(z), z)}{s(f(a), a)}.$$

This inequality, together with (2.2.4), give the inequality of Theorem 2.5, completing the proof.

2.3. Holomorphic maps of the disc

This section proves Theorem 2.1. The following is a preliminary lemma about conformal automorphisms of the unit disc.

LEMMA 2.8. *Let c be a point that lies on the axis of a hyperbolic automorphism h of \mathbb{D} . Then*

$$\rho(w, h(w)) \leq e^{\rho(w, c)} \rho(c, h(c)),$$

for all $w \in \mathbb{D}$.

PROOF. Let γ be the axis of h . By [6, Theorem 7.35.1], we have

$$\sinh \frac{1}{2} \rho(w, h(w)) = \cosh \rho(w, \gamma) \sinh \frac{1}{2} \rho(c, h(c)),$$

for $c \in \gamma$ and $w \in \mathbb{D}$. Now, as we mentioned earlier, the function $x \mapsto \sinh x/x$ is increasing for $x > 0$, and $\rho(c, h(c)) \leq \rho(w, h(w))$, so

$$\frac{\rho(w, h(w))}{\rho(c, h(c))} \leq \frac{\sinh \frac{1}{2} \rho(w, h(w))}{\sinh \frac{1}{2} \rho(c, h(c))} = \cosh \rho(w, \gamma) \leq e^{\rho(w, \gamma)}.$$

The result then follows from the inequality $\rho(w, \gamma) \leq \rho(w, c)$. \square

Now we prove Theorem 2.1. Suppose, then, that f is a holomorphic self-map of \mathbb{D} and that a, b and z are points in \mathbb{D} , with $a \neq b$. If f does not fix b , then there is a unique hyperbolic line γ through b and $f(b)$. Let h be the hyperbolic automorphism of \mathbb{D} with axis γ that satisfies $hf(b) = b$. If f fixes b , then we define h to be the identity function. So, applying Theorem 2.5 to hf , we see that

$$\rho(hf(z), z) \leq M \rho(hf(a), a), \quad \text{where} \quad M = \frac{e^{\rho(a, z) + \rho(z, b)}}{4 \sinh \frac{1}{2} \rho(a, b)}.$$

Hence

$$\begin{aligned} \rho(f(z), z) &\leq \rho(f(z), hf(z)) + \rho(hf(z), z) \\ &\leq \rho(f(z), hf(z)) + M \rho(hf(a), a) \\ &\leq \rho(f(z), hf(z)) + M \rho(f(a), hf(a)) + M \rho(f(a), a). \end{aligned}$$

Next, for $u \in \mathbb{D}$, Lemma 2.8 (with $w = f(u)$ and $c = f(b)$) tells us that if $f(b) \neq b$, then

$$\rho(f(u), hf(u)) \leq e^{\rho(f(u), f(b))} \rho(f(b), hf(b)) \leq e^{\rho(u, b)} \rho(f(b), b),$$

using the Schwarz–Pick lemma for the final inequality. Clearly, this inequality also holds if $f(b) = b$. Since $e^{\rho(a,b)} > 4 \sinh \frac{1}{2}\rho(a,b)$, we see that

$$\begin{aligned}
 \rho(f(z), z) &\leq e^{\rho(z,b)} \rho(f(b), b) + M e^{\rho(a,b)} \rho(f(b), b) + M \rho(f(a), a) \\
 &\leq (e^{\rho(z,b)} + M e^{\rho(a,b)}) (\rho(f(a), a) + \rho(f(b), b)) \\
 (2.3.1) \quad &\leq \frac{e^{\rho(a,z) + \rho(a,b) + \rho(z,b)}}{2 \sinh \frac{1}{2}\rho(a,b)} (\rho(f(a), a) + \rho(f(b), b)).
 \end{aligned}$$

Theorem 2.1 can be deduced from (2.3.1), since $\sinh x \geq x$ for all $x > 0$. However, we highlight the slightly stronger inequality of (2.3.1) for use later.

2.4. Holomorphic maps of the punctured disc

It remains to prove Theorem 2.3, and this section is dedicated to that task. Recall that $\lambda^*(z)|dz| = -|dz|/(|z| \log |z|)$ is the metric associated with the hyperbolic distance on the punctured unit disc. We use the following trivial estimates.

LEMMA 2.9. *If $z \in \mathbb{D}^*$, then*

- (1) $\lambda^*(z) \geq e$,
- (2) $\lambda^*(z) \geq -\log |z|$,
- (3) $\lambda^*(z) \geq -\frac{1}{\log |z|}$.

Let us suppose, as stated in Theorem 2.3, that f is a holomorphic self-map of \mathbb{D}^* , h is a holomorphic self-covering map of \mathbb{D}^* , and $\deg(f)$ and $\deg(h)$ are both equal to some positive integer m . By post-composing f and h with a suitable rotation of \mathbb{D}^* about 0 (a hyperbolic isometry of \mathbb{D}^*) we can assume that $h(z) = z^m$.

Suppose that $a, z \in \mathbb{D}^*$. Let $\pi: \mathbb{H} \rightarrow \mathbb{D}^*$ be the universal covering map $\pi(\zeta) = e^{2\pi i \zeta}$. We denote the hyperbolic metric on \mathbb{H} by ρ . Let \tilde{a} be any point in \mathbb{H} such that $\pi(\tilde{a}) = a$. Since

$$\rho^*(z, a) = \inf\{\rho(\zeta, \tilde{a}) : \zeta \in \mathbb{H} \text{ and } \pi(\zeta) = z\},$$

we can choose $\tilde{z} \in \mathbb{H}$ such that $\pi(\tilde{z}) = z$ and $\rho(\tilde{z}, \tilde{a}) = \rho^*(z, a)$ (see [1, Section 1.1.4]).

We now lift the map f to a holomorphic map $\tilde{f}: \mathbb{H} \rightarrow \mathbb{H}$ with $\pi \circ \tilde{f} = f \circ \pi$. The map \tilde{f} satisfies $\tilde{f}(\zeta + 1) = \tilde{f}(\zeta) + m$, for all $\zeta \in \mathbb{H}$, since f has degree m . We also lift the map

h to the holomorphic map $\tilde{h}(\zeta) = m\zeta$, which is a hyperbolic isometry of \mathbb{H} . Thus the two diagrams below commute.

$$\begin{array}{ccc} \mathbb{H} & \xrightarrow{\tilde{f}} & \mathbb{H} \\ \pi \downarrow & & \downarrow \pi \\ \mathbb{D}^* & \xrightarrow{f} & \mathbb{D}^* \end{array} \qquad \begin{array}{ccc} \mathbb{H} & \xrightarrow{\tilde{h}} & \mathbb{H} \\ \pi \downarrow & & \downarrow \pi \\ \mathbb{D}^* & \xrightarrow{h} & \mathbb{D}^* \end{array}$$

Observe that $\pi(\tilde{f}(\tilde{a})) = f(a)$ and $\pi(\tilde{h}(\tilde{a})) = h(a)$. By replacing \tilde{f} with a map that is the composition of \tilde{f} followed by a suitable integer translation (also a lift of f), we can assume that $\rho(\tilde{f}(\tilde{a}), \tilde{h}(\tilde{a})) = \rho^*(f(a), h(a))$. Next we apply the slightly stronger version of Theorem 2.1 given by inequality (2.3.1) to the function $\tilde{h}^{-1} \circ \tilde{f}$ and the points \tilde{a} , $\tilde{a} + 1$ and \tilde{z} . We obtain

$$\rho(\tilde{h}^{-1} \circ \tilde{f}(\tilde{z}), \tilde{z}) \leq K(\rho(\tilde{h}^{-1} \circ \tilde{f}(\tilde{a}), \tilde{a}) + \rho(\tilde{h}^{-1} \circ \tilde{f}(\tilde{a} + 1), \tilde{a} + 1)),$$

where $K = e^{\rho(\tilde{z}, \tilde{a}) + \rho(\tilde{a}, \tilde{a} + 1) + \rho(\tilde{a} + 1, \tilde{z})} / (2 \sinh \frac{1}{2} \rho(\tilde{a}, \tilde{a} + 1))$. Since \tilde{h} is a hyperbolic isometry of \mathbb{H} , we see that

$$\begin{aligned} \rho(\tilde{f}(\tilde{z}), \tilde{h}(\tilde{z})) &\leq K(\rho(\tilde{f}(\tilde{a}), \tilde{h}(\tilde{a})) + \rho(\tilde{f}(\tilde{a} + 1), \tilde{h}(\tilde{a} + 1))) \\ &= K(\rho(\tilde{f}(\tilde{a}), \tilde{h}(\tilde{a})) + \rho(\tilde{f}(\tilde{a}) + m, \tilde{h}(\tilde{a}) + m)) \\ &= 2K\rho(\tilde{f}(\tilde{a}), \tilde{h}(\tilde{a})). \end{aligned}$$

Now

$$\begin{aligned} K &= \frac{\exp(\rho(\tilde{z}, \tilde{a}) + \rho(\tilde{a}, \tilde{a} + 1) + \rho(\tilde{a} + 1, \tilde{z}))}{2 \sinh \frac{1}{2} \rho(\tilde{a}, \tilde{a} + 1)} \\ &\leq \frac{\exp(2\rho(\tilde{z}, \tilde{a}) + 2\rho(\tilde{a}, \tilde{a} + 1))}{2 \sinh \frac{1}{2} \rho(\tilde{a}, \tilde{a} + 1)} \\ &\leq 2e^{2\rho(\tilde{z}, \tilde{a})} \frac{\cosh^2 \rho(\tilde{a}, \tilde{a} + 1)}{\sinh \frac{1}{2} \rho(\tilde{a}, \tilde{a} + 1)}. \end{aligned}$$

We can write this expression in Euclidean terms by using the following standard formulas for the hyperbolic metric on \mathbb{H} , taken from [10, Theorem 7.4]:

$$\cosh \rho(u, v) = 1 + \frac{|u - v|^2}{2 \operatorname{Im} u \operatorname{Im} v}, \quad \sinh \frac{1}{2} \rho(u, v) = \frac{|u - v|}{2\sqrt{\operatorname{Im} u \operatorname{Im} v}},$$

where $u, v \in \mathbb{H}$. Thus

$$K \leq 2e^{2\rho(\tilde{z}, \tilde{a})} \frac{(1 + \frac{1}{2}(\operatorname{Im} \tilde{a})^{-2})^2}{\frac{1}{2}(\operatorname{Im} \tilde{a})^{-1}} = e^{2\rho(\tilde{z}, \tilde{a})} (4 \operatorname{Im} \tilde{a} + 4(\operatorname{Im} \tilde{a})^{-1} + (\operatorname{Im} \tilde{a})^{-3}).$$

Now $e^{2\pi i \tilde{a}} = a$, so $e^{-2\pi \operatorname{Im} \tilde{a}} = |a|$. Hence $\operatorname{Im} \tilde{a} = -(\log |a|)/(2\pi)$, and we can apply Lemma 2.9 to deduce that $\operatorname{Im} \tilde{a} \leq \lambda^*(a)/(2\pi)$ and $(\operatorname{Im} \tilde{a})^{-1} \leq 2\pi \lambda^*(a)$. Since $\lambda^*(a) \geq e$, we obtain

$$4 \operatorname{Im} \tilde{a} + 4(\operatorname{Im} \tilde{a})^{-1} + (\operatorname{Im} \tilde{a})^{-3} \leq (1 + 8\pi) \lambda^*(a) + (2\pi)^3 \lambda^*(a)^3 \leq 252 \lambda^*(a)^3.$$

Therefore

$$\rho(\tilde{f}(\tilde{z}), \tilde{h}(\tilde{z})) \leq 504 \lambda^*(a)^3 e^{2\rho(\tilde{z}, \tilde{a})} \rho(\tilde{f}(\tilde{a}), \tilde{h}(\tilde{a})).$$

The proof of Theorem 2.3 is complete on observing that

$$\rho^*(f(z), h(z)) \leq \rho(\tilde{f}(\tilde{z}), \tilde{h}(\tilde{z})), \quad \rho^*(f(a), h(a)) = \rho(\tilde{f}(\tilde{a}), \tilde{h}(\tilde{a})) \quad \text{and} \quad \rho^*(z, a) = \rho(\tilde{z}, \tilde{a}).$$

2.5. Theorem 2.1 in geometric function theory

In this section we describe how our main result from this chapter, Theorem 2.1, can provide an alternative proof of the following rigidity theorem due to Burns and Kranz [15, Theorem 2.1].

THEOREM 2.10 ([15]). *Suppose that f is holomorphic self-map of \mathbb{D} such that*

$$f(z) = z + o(|z - 1|^3),$$

as $z \rightarrow 1$. Then f is the identity function.

The motivation behind this theorem comes from a generalisation of Schwarz's lemma proved by H. Cartan in 1931 [38, page 66]. This result, often called Cartan's uniqueness theorem, states that if a function $f \in \operatorname{Hol}(\mathbb{D})$ is such that $f(z) = z + o(|z|)$, as $z \rightarrow 0$, then f is the identity function. So, Theorem 2.10 can be seen as a generalisation of Cartan's theorem where the origin is replaced by the boundary point 1, which is the reason it is

often described as a “boundary Schwarz lemma”. The point 1 in Theorem 2.10 and the origin in Cartan’s uniqueness theorem can of course be replaced with any points $\zeta \in \partial\mathbb{D}$ and $z_0 \in \mathbb{D}$, respectively. What is also interesting about Theorem 2.10, is the fact that the exponent 3 is sharp, as one can verify from the function

$$g(z) = z - \frac{(z-1)^3}{10}.$$

Both results, by Cartan and Burns, Krantz, were originally stated for bounded convex domains in \mathbb{C}^n , with suitably smooth boundary.

Recently, our result, Theorem 2.1, was used by Zimmer [47, Theorem 1.5] in order to generalise Theorem 2.10 to a larger class of domains in \mathbb{C}^n . Zimmer [47, Section 4] also showed that it is possible to obtain Theorem 2.10 using Theorem 2.1 and elementary techniques in hyperbolic geometry. This, in conjunction with our proof of Theorem 2.1, shows that it is possible to obtain Theorem 2.10 of Burns, Krantz using only the hyperbolic geometry of the domain \mathbb{D} , instead of the deeper potential theoretic techniques originally used in [15].

CHAPTER 3

Stability of the Denjoy–Wolff theorem

3.1. Introduction and background

Motivated by the goal of generalising the classical Denjoy–Wolff theorem, Theorem 1.3, we examine the stability of the Denjoy–Wolff theorem under perturbations of the holomorphic map f , in a sense to be made precise shortly.

Recall that $\text{Hol}(\mathbb{D})$ is the topological space of all holomorphic self-maps of \mathbb{D} , equipped with the compact-open topology. If (f_n) is a sequence in $\text{Hol}(\mathbb{D})$ that converges locally uniformly on \mathbb{D} to a function f , then either $f \in \text{Hol}(\mathbb{D})$ or else f is a constant function with value on the boundary of \mathbb{D} (see [8, Lemma 2.1]).

Given sequences (f_n) and (g_n) in $\text{Hol}(\mathbb{D})$, we define the *left-composition sequence* generated by (f_n) and the *right-composition sequence* generated by (g_n) to be the sequences

$$F_n = f_n \circ f_{n-1} \circ \cdots \circ f_1 \quad \text{and} \quad G_n = g_1 \circ g_2 \circ \cdots \circ g_n, \quad n = 1, 2, \dots,$$

respectively. Sequences of this type arise in a variety of contexts in dynamical systems, with differing notations and terminology. In all later chapters we omit the \circ symbol from compositions.

The dynamical behaviour of the sequence of iterates (f^n) , where $f \in \text{Hol}(\mathbb{D})$, depends on whether f is the identity function, an elliptic Möbius transformation, or if it has a Denjoy–Wolff point that lies in \mathbb{D} or on the boundary of \mathbb{D} . We determine whether the dynamics of (F_n) and (G_n) are similar to that of (f^n) under the assumptions that $f_n \rightarrow f$ and $g_n \rightarrow f$. We find that, in a sense, right-composition sequences are more stable than left-composition sequences when f has a Denjoy–Wolff point inside \mathbb{D} , but the reverse holds when the Denjoy–Wolff point of f lies on the boundary of \mathbb{D} . And when f is the identity function, there is similar stability for both left- and right-composition sequences.

We will make significant use of the hyperbolic metric on \mathbb{D} and the Schwarz–Pick lemma. Note that if f is not a conformal automorphism, then for each compact subset K of \mathbb{D} we can find a positive constant $k < 1$ such that $\rho(f(z), f(w)) \leq k\rho(z, w)$, for $z, w \in K$, because f is a strict contraction of the hyperbolic metric in \mathbb{D} .

There are various generalisations of the Denjoy–Wolff theorem in the literature, and we now give a brief outline of the previous results most relevant to our analysis. One of the first theorems in this area was the following theorem from [23, Theorem 1], [4, Corollary 2.3] and [33, Theorem 1.2].

THEOREM 3.1. *Suppose that K is a compact subset of a simply connected hyperbolic domain D , and that g_1, g_2, \dots are holomorphic maps of D into K . Then the right-composition sequence $G_n = g_1 g_2 \cdots g_n$ converges locally uniformly on D to a constant in K .*

Other noteworthy generalisations of the Denjoy–Wolff theorem similar to Theorem 3.1 can be found in [8, 27, 32]. A slightly different approach to the subject was considered by Beardon, Carne, Minda and Ng [9], where instead of examining the convergence of a right-composition sequence $f_1 f_2 \dots$ directly, they study the hyperbolic density of the images of \mathbb{D} under each f_i for $i = 1, 2, \dots$. In particular, they prove that if each f_i exhibits strong contracting properties with respect to the hyperbolic metric, then the composition sequence $f_1 f_2 \dots$ can only have constant limit functions. For more information on the hyperbolic density of hyperbolic domains, as well as results similar to the Beardon, Carne, Minda, Ng, we refer to [29], [30, Chapters 11 & 12] and [45].

There is an extensive literature on stability results for holomorphic dynamical systems; we draw attention to the papers of Beardon [7], Gill [23, 24] and Pommerenke [41] for work closest to our own. Beardon and Gill were motivated in part by the theory of limit-periodic continued fractions, in which one considers the stability of continued fractions under perturbations of the coefficients. In [7], Beardon looks at the stability of Möbius transformations under iteration. We develop the geometric approach of [7], and apply it to the class of holomorphic maps, which is far larger and more complex than the class of Möbius transformations. Note that Theorem 3.5 of Section 3.3 could be deduced quickly from [7, Theorem 4.7] (the proof we give is short anyway).

Gill studies composition sequences of holomorphic maps for which the constituent maps approach a limit function. Using Euclidean estimates he obtains results of a similar type to Theorems 3.5 and 3.7. One of the benefits of our geometric approach is that we obtain strong results with succinct statements and concise proofs using the hyperbolic metric.

Pommerenke considers right-composition sequences (F_n) under the assumption that $f_n \rightarrow f$, for some non-elliptic map f , and attempts to find constants a_n and b_n such that $a_n F_n + b_n \rightarrow F$, for some non-constant function F . Whether this is possible depends on the nature of the Denjoy–Wolff point of f . Our objectives are somewhat tangential to this, such that we obtain a complete analysis of stability for both left- and right- composition sequences and any choice of holomorphic map f .

3.2. Stability at elliptic transformations and the identity function

Here we consider the behaviour of the left- and right-composition sequences $F_n = f_n f_{n-1} \cdots f_1$ and $G_n = g_1 g_2 \cdots g_n$, where $f_n, g_n \in \text{Hol}(\mathbb{D})$, under the assumption that the sequences (f_n) and (g_n) converge to an elliptic Möbius transformation fixing \mathbb{D} or the identity function Id . We focus particularly on the latter case, because the iterates of an elliptic transformation do not themselves converge in $\text{Hol}(\mathbb{D})$.

The next example demonstrates that when $f_n \rightarrow \text{Id}$, and without further assumptions, the sequence (F_n) can behave erratically.

EXAMPLE 3.2. Let $f_n(z) = e^{i/n}z$, for $n = 1, 2, \dots$, so $f_n \rightarrow \text{Id}$. Then

$$F_n(z) = e^{i\left(1 + \frac{1}{2} + \cdots + \frac{1}{n}\right)} z.$$

This sequence accumulates at the identity function and every rotation of the unit circle. \square

Essentially the same example can be used with g_n in place of f_n and G_n in place of F_n , because the functions commute.

We can get quite different behaviour with other choices for functions $f_n \rightarrow \text{Id}$. For example, choosing $f_n(z) = (1 - 1/n)z$, for $n = 2, 3, \dots$, we see that (F_n) converges locally uniformly on \mathbb{D} to 0.

Example 3.2 indicates that to obtain more controlled behaviour of (F_n) and (G_n) under the assumption that $f_n \rightarrow \text{Id}$ and $g_n \rightarrow \text{Id}$ we need additional constraints on convergence. Theorems 3.3 and 3.4, to follow, show that such control can be achieved if we stipulate that the convergence is sufficiently fast (in a sense to be made precise). In fact, using Theorem 2.1 from Chapter 2, we will see that it is sufficient to assume that (f_n) and (g_n) converge to the identity function suitably fast at just two points in \mathbb{D} .

We now state our first result about stability of the Denjoy–Wolff theorem at the identity function or an elliptic transformation, for left-composition sequences.

THEOREM 3.3. *Suppose that f is either the identity function or an elliptic Möbius transformation that fixes \mathbb{D} , and f_1, f_2, \dots are non-constant holomorphic self-maps of \mathbb{D} for which*

$$\sum_{n=1}^{\infty} \rho(f_n(a), f(a)) < +\infty \quad \text{and} \quad \sum_{n=1}^{\infty} \rho(f_n(b), f(b)) < +\infty,$$

for two distinct points $a, b \in \mathbb{D}$. Then the sequence $(f^{-n}F_n)$, where $F_n = f_n f_{n-1} \cdots f_1$, converges locally uniformly on \mathbb{D} to a non-constant holomorphic self-map of \mathbb{D} .

PROOF. Let $d = \frac{1}{3}\rho(a, b)$ and let K be a closed hyperbolic disc that is centred at a fixed point of f and contains a and b . Observe that if $z \in K$, then $f^n(z) \in K$, for $n \in \mathbb{Z}$. By applying Corollary 2.2 to the functions f_n and f , for $n = 1, 2, \dots$, we see that

$$\sum_{n=1}^{\infty} \sup_{z \in K} \rho(f_n(z), f(z)) < +\infty.$$

Notice that it suffices to prove the theorem for the truncated left-composition sequence with n th term $f_n f_{n-1} \cdots f_N$, where N is a fixed positive integer. In light of this observation, we may assume (after relabelling the functions) that in fact

$$\sum_{n=1}^{\infty} \sup_{z \in K} \rho(f_n(z), f(z)) < d.$$

Choose any point $z \in K$. Let $z_n = f^n(z)$, for $n = 1, 2, \dots$. Then $z_n \in K$. Observe that

$$\begin{aligned} \rho(F_n(z), f^n(z)) &\leq \rho(f_n \cdots f_1(z), f_n \cdots f_2 f(z)) + \rho(f_n \cdots f_2(f(z)), f^{n-1}(f(z))) \\ &\leq \rho(f_1(z), f(z)) + \rho(f_n \cdots f_2(z_1), f^{n-1}(z_1)), \end{aligned}$$

where, to obtain the second inequality, we have applied the Schwarz–Pick lemma with the function $f_n \cdots f_2$. Repeating this argument we see that

$$\rho(F_n(z), f^n(z)) \leq \rho(f_1(z), f(z)) + \rho(f_2(z_1), f(z_1)) + \cdots + \rho(f_n(z_{n-1}), f(z_{n-1})) < d,$$

for $n = 1, 2, \dots$.

Next, still with $z \in K$, we have

$$\begin{aligned} \rho(F_n(z), a) &\leq \rho(F_n(z), F_n(a)) + \rho(F_n(a), f^n(a)) + \rho(f^n(a), a) \\ &\leq \rho(z, a) + d + \rho(f^n(a), a) \leq l, \end{aligned}$$

for $n = 1, 2, \dots$, where l is three times the hyperbolic diameter of K . Similarly $\rho(F_n(z), b) \leq l$. Applying Corollary 2.2 to the functions f_n and f , and with $F_{n-1}(z)$ in place of z , we obtain

$$\rho(F_n(z), f(F_{n-1}(z))) \leq \lambda(\rho(f_n(a), f(a)) + \rho(f_n(b), f(b))),$$

where

$$\lambda = \frac{\exp(\rho(F_{n-1}(z), a) + \rho(a, b) + \rho(b, F_{n-1}(z)))}{\rho(a, b)} \leq \frac{\exp(3l)}{\rho(a, b)}.$$

Consequently, we see that

$$\sum_{n=1}^{\infty} \rho(f^{-n}F_n(z), f^{-(n-1)}(F_{n-1}(z))) = \sum_{n=1}^{\infty} \rho(F_n(z), f(F_{n-1}(z))) < 2\lambda d,$$

for $z \in K$ (where F_0 is the identity function). Thus $(f^{-n}F_n)$ is a uniformly Cauchy sequence on K . Now, K is an arbitrarily large compact subset of \mathbb{D} , so it follows that $(f^{-n}F_n)$ converges locally uniformly on \mathbb{D} to a function F .

The function F belongs to $\text{Hol}(\mathbb{D})$, and it is not a constant function because

$$\rho(f^{-n}F_n(a), f^{-n}F_n(b)) \geq \rho(a, b) - \rho(f^{-n}F_n(a), a) - \rho(f^{-n}F_n(b), b) > 3d - d - d = d,$$

for $n = 1, 2, \dots$. □

When f is the identity function I , Theorem 3.3 says that if $\sum \rho(f_n(a), a) < +\infty$ and $\sum \rho(f_n(b), b) < +\infty$, then the left-composition sequence $F_n = f_n f_{n-1} \cdots f_1$ converges locally uniformly on \mathbb{D} to a non-constant holomorphic map $F \in \text{Hol}(\mathbb{D})$. And when f is an elliptic transformation of finite order m , the theorem tells us that the sequence (F_n) can be split into m subsequences that converge to $F, fF, \dots, f^{m-1}F$, respectively. For

the remaining case, when f is an elliptic transformation of infinite order, we see from Theorem 3.3 that (F_n) accumulates at uncountably many different non-constant maps in $\text{Hol}(\mathbb{D})$.

Next we state a result similar to Theorem 3.3 for right-composition sequences.

THEOREM 3.4. *Suppose that g is either the identity function or an elliptic Möbius transformation that fixes \mathbb{D} , and g_1, g_2, \dots are non-constant holomorphic self-maps of \mathbb{D} for which*

$$\sum_{n=1}^{\infty} \rho(g_n(a), g(a)) < +\infty \quad \text{and} \quad \sum_{n=1}^{\infty} \rho(g_n(b), g(b)) < +\infty,$$

for two distinct points $a, b \in \mathbb{D}$. Then the sequence $(G_n g^{-n})$, where $G_n = g_1 g_2 \cdots g_n$, converges locally uniformly on \mathbb{D} to a non-constant holomorphic self-map of \mathbb{D} .

PROOF. Let $d = \frac{1}{3}\rho(a, b)$ and let K be a closed hyperbolic disc that is centred at a fixed point of g and that contains a and b . By truncating the right-composition sequence (G_n) by a fixed finite number of terms from the left (and relabelling the remaining functions), we can assume that

$$\sum_{n=1}^{\infty} \sup_{z \in K} \rho(g_n(z), g(z)) < d.$$

Now choose a point z in K , and let n be a positive integer. By applying the Schwarz–Pick lemma with the function G_{n-1} , we see that

$$\rho(G_n g^{-n}(z), G_{n-1} g^{-(n-1)}(z)) \leq \rho(g_n(w), g(w)),$$

where $w = g^{-n}(z)$ (and G_0 is the identity function). Since $w \in K$, it follows that

$$\sum_{n=1}^{\infty} \rho(G_n g^{-n}(z), G_{n-1} g^{-(n-1)}(z)) < d.$$

Therefore $(G_n g^{-n})$ is a uniformly Cauchy sequence on K , and since K can be chosen to be arbitrarily large, we deduce that $(G_n g^{-n})$ converges locally uniformly on \mathbb{D} to a function G .

This function G belongs to $\text{Hol}(\mathbb{D})$; we must show that it is not a constant function. To this end, we write $a_n = g^{-n}(a)$, for $n = 1, 2, \dots$, and observe that

$$\begin{aligned} \rho(G_n g^{-n}(a), a) &\leq \rho(G_n(a_n), G_{n-1}(a_{n-1})) + \rho(G_{n-1}(a_{n-1}), G_{n-2}(a_{n-2})) + \dots + \rho(G_1(a_1), a) \\ &\leq \rho(g_n(a_n), g(a_n)) + \rho(g_{n-1}(a_{n-1}), g(a_{n-1})) + \dots + \rho(g_1(a_1), g(a_1)), \end{aligned}$$

for $n = 1, 2, \dots$, where, to obtain the second inequality, we applied the Schwarz–Pick lemma with the functions $G_{n-1}, G_{n-2}, \dots, G_0$, in that order. Since $a_n \in K$, for each index n , we find that $\rho(G_n g^{-n}(a), a) < d$, and similarly $\rho(G_n g^{-n}(b), b) < d$. Consequently,

$$\rho(G_n g^{-n}(a), G_n g^{-n}(b)) \geq \rho(a, b) - \rho(G_n g^{-n}(a), a) - \rho(G_n g^{-n}(b), b) > 3d - d - d = d,$$

for $n = 1, 2, \dots$. Hence G is a non-constant holomorphic self-map of \mathbb{D} . \square

The special cases of Theorem 3.4 when the limit function g is of finite order resemble the similar special cases of Theorem 3.3. In particular, when g is the identity function, Theorem 3.4 says that if $\sum \rho(g_n(a), a) < +\infty$ and $\sum \rho(g_n(b), b) < +\infty$, then the right-composition sequence $G_n = g_1 g_2 \dots g_n$ converges locally uniformly on \mathbb{D} to a non-constant holomorphic self-map of \mathbb{D} .

3.3. Denjoy–Wolff point inside the disc

In this section we consider the stability of the Denjoy–Wolff theorem at holomorphic functions that have a Denjoy–Wolff point inside the unit disc. Using Theorem 3.1 we obtain the following strong stability result for right-composition sequences.

THEOREM 3.5. *Let g be a holomorphic self-map of \mathbb{D} with a Denjoy–Wolff point ζ in \mathbb{D} . Then there is a neighbourhood \mathcal{U} of g in $\text{Hol}(\mathbb{D})$ such that if g_1, g_2, \dots belong to \mathcal{U} , then the right-composition sequence $G_n = g_1 g_2 \dots g_n$ converges locally uniformly on \mathbb{D} to a constant in \mathbb{D} .*

We use the notation $D(c, r)$ for the hyperbolic open disc with centre c and radius r .

PROOF. Let $D = D(\zeta, r)$, for some $r > 0$. Since \overline{D} is a compact set in \mathbb{D} , we see from the Schwarz–Pick lemma that there is a positive constant $k < 1$ (that depends on \overline{D}) with

$\rho(g(z), g(w)) \leq k\rho(z, w)$, for $z, w \in \overline{D}$. Observe that g fixes ζ , so $g(D) \subset D(\zeta, s)$, where $s = kr$. Now choose a real number t with $s < t < r$. Let

$$\mathcal{U} = \{h \in \text{Hol}(\mathbb{D}) : h(D) \subset D(\zeta, t)\},$$

a neighbourhood of g in $\text{Hol}(\mathbb{D})$, and let $K = \overline{D(\zeta, t)}$. If g_1, g_2, \dots belong to \mathcal{U} , then $g_n(D) \subset K$, for each index n , so we can apply Theorem 3.1 to see that the right-composition sequence $G_n = g_1 g_2 \cdots g_n$ converges locally uniformly on D to a constant in K . Since the set D has an accumulation point in \mathbb{D} , the Vitali-Porter Theorem (see, for example, [43, Section 2.4]) yields that (G_n) converges locally uniformly on \mathbb{D} to a constant in \mathbb{D} . \square

The hypotheses of Theorem 3.5 can of course be weakened to assume that all but finitely many of the maps g_n belong to \mathcal{U} .

The next example shows that there is no analogue of Theorem 3.5 for left-composition sequences.

EXAMPLE 3.6. Let $f(z) = z/2$, and let \mathcal{U} be a neighbourhood of f in $\text{Hol}(\mathbb{D})$. We can choose a positive constant δ sufficiently small that all the functions $f_n(z) = z/2 + \delta e^{i\theta_n}$, where $\theta_n \in \mathbb{R}$, for $n = 1, 2, \dots$, belong to \mathcal{U} . The left-composition sequence $F_n = f_n f_{n-1} \cdots f_1$ satisfies

$$F_n(z) = \frac{1}{2}F_{n-1}(z) + \delta e^{i\theta_n}.$$

Evidently, the parameters θ_n can be chosen so that (F_n) diverges pointwise on \mathbb{D} . \square

With slightly stronger hypotheses, however, we do obtain controlled behaviour of the left-composition sequence (F_n) .

THEOREM 3.7. *Let f be a holomorphic self-map of \mathbb{D} with a Denjoy–Wolff point ζ in \mathbb{D} . Suppose that f_1, f_2, \dots is a sequence of functions in $\text{Hol}(\mathbb{D})$ that converges locally uniformly on \mathbb{D} to f . Then the left-composition sequence $F_n = f_n f_{n-1} \cdots f_1$ converges locally uniformly on \mathbb{D} to ζ .*

PROOF. Let K be a closed hyperbolic disc centred at ζ . Observe that f maps K inside a smaller closed hyperbolic disc centred at ζ . Since $f_n \rightarrow f$ uniformly on K we see that

f_n maps K inside itself for sufficiently large n . By truncating F_n by finitely many terms on the right (and relabelling) we can assume that in fact $f_n(K) \subset K$ for all $n = 1, 2, \dots$.

We define k to be a constant between 0 and 1 for which $\rho(f(z), f(w)) \leq k\rho(z, w)$, for $z, w \in K$.

Choose $z \in K$. Observe that $f^n(z) \in K$ and $F_n(z) \in K$, for $n = 1, 2, \dots$. Then

$$\begin{aligned} \rho(F_n(z), f^n(z)) &\leq \rho(F_n(z), f(F_{n-1}(z))) + \rho(f(F_{n-1}(z)), f^n(z)) \\ &\leq \sup_{w \in K} \rho(f_n(w), f(w)) + k\rho(F_{n-1}(z), f^{n-1}(z)), \end{aligned}$$

for $n = 1, 2, \dots$. Iterating this argument, we see that

$$\rho(F_n(z), f^n(z)) \leq (1 + k + k^2 + \dots + k^{n-1}) \sup_{w \in K} \rho(f_n(w), f(w)) \leq \frac{1}{1 - k} \sup_{w \in K} \rho(f_n(w), f(w)),$$

for $n = 1, 2, \dots$. Since (f_n) converges locally uniformly on \mathbb{D} to f we see that $\rho(F_n(z), f^n(z)) \rightarrow 0$ uniformly on K , so $F_n \rightarrow \zeta$ uniformly on K . Hence (F_n) converges locally uniformly on \mathbb{D} to the constant ζ . \square

Notice that the left-composition sequence (F_n) of Theorem 3.7 converges locally uniformly on \mathbb{D} to ζ , but the right-composition sequence (G_n) of Theorem 3.5 converges to a constant that need not be ζ . This is because adjusting g_1 causes the constant to change.

3.4. Denjoy–Wolff point on the boundary of the disc

This final section considers the stability of the Denjoy–Wolff theorem at holomorphic functions f that have a Denjoy–Wolff point on the boundary of the unit disc. In a sense, this circumstance is the least stable of those considered so far. Indeed, it is straightforward to find holomorphic maps f_1, f_2, \dots with $f_n \rightarrow f$ (for a suitable choice of f with a Denjoy–Wolff point on the boundary of \mathbb{D}) for which the behaviour of the left-composition sequence $F_n = f_n f_{n-1} \dots f_1$ is erratic. Nevertheless, the following theorem shows that if we assume that the convergence of (f_n) to f is sufficiently rapid, then the sequences (F_n) and (f^n) have similar dynamics.

THEOREM 3.8. *Let f be a holomorphic self-map of \mathbb{D} with a Denjoy–Wolff point ζ on the boundary of \mathbb{D} . Then there exist neighbourhoods $\mathcal{U}_1, \mathcal{U}_2, \dots$ of f in $\text{Hol}(\mathbb{D})$ such that if*

$f_n \in \mathcal{U}_n$, for $n = 1, 2, \dots$, then the left-composition sequence $F_n = f_n f_{n-1} \cdots f_1$ converges locally uniformly on \mathbb{D} to ζ .

PROOF. For each positive integer n , we define D_n to be the open hyperbolic disc centred at 0 of radius $1 + \rho(f^{n-1}(0), 0)$, and let

$$\mathcal{U}_n = \{h \in \text{Hol}(\mathbb{D}) : \rho(h(z), f(z)) < 1/2^n \text{ for } z \in D_n\},$$

a neighbourhood of f in $\text{Hol}(\mathbb{D})$. Suppose that $f_n \in \mathcal{U}_n$, for $n = 1, 2, \dots$.

We will prove by induction on m that

$$\rho(F_m(0), f^m(0)) < 1 - \frac{1}{2^m},$$

for $m = 1, 2, \dots$. This is certainly true for $m = 1$, by definition of \mathcal{U}_1 . Suppose that it is true for the integer $m = n - 1$, where $n > 1$. Then

$$\begin{aligned} \rho(F_n(0), f^n(0)) &\leq \rho(F_n(0), f(F_{n-1}(0))) + \rho(f(F_{n-1}(0)), f^n(0)) \\ &\leq \rho(F_n(0), f(F_{n-1}(0))) + \rho(F_{n-1}(0), f^{n-1}(0)) \\ &< \rho(F_n(0), f(F_{n-1}(0))) + 1 - \frac{1}{2^{n-1}}, \end{aligned}$$

where we have applied the triangle inequality, the Schwarz–Pick lemma, and the induction hypothesis. Now, since

$$\rho(F_{n-1}(0), 0) \leq \rho(F_{n-1}(0), f^{n-1}(0)) + \rho(f^{n-1}(0), 0) < 1 + \rho(f^{n-1}(0), 0),$$

we see that $F_{n-1}(0) \in D_n$. So, by definition of \mathcal{U}_n , we have

$$\rho(F_n(0), f(F_{n-1}(0))) = \rho(f_n(F_{n-1}(0)), f(F_{n-1}(0))) < \frac{1}{2^n}.$$

Combining the inequalities obtained we conclude that

$$\rho(F_n(0), f^n(0)) < \rho(F_n(0), f(F_{n-1}(0))) + 1 - \frac{1}{2^{n-1}} < \frac{1}{2^n} + 1 - \frac{1}{2^{n-1}} = 1 - \frac{1}{2^n}.$$

This completes the proof by induction.

A consequence of this observation is that $\rho(F_n(0), f^n(0)) < 1$, for each positive integer n . Then, since $f^n(0) \rightarrow \zeta$, a point on the boundary of \mathbb{D} , we can use a formula for the

hyperbolic metric in \mathbb{D} such as

$$\sinh \frac{1}{2}\rho(z, w) = \frac{|z - w|}{\sqrt{(1 - |z|^2)(1 - |w|^2)}},$$

to see that $F_n(0) \rightarrow \zeta$ also.

Furthermore, we have that $\rho(F_n(z), F_n(0)) \leq \rho(z, 0)$, for any point $z \in \mathbb{D}$, and from this inequality we see that (F_n) converges locally uniformly on \mathbb{D} to ζ (with convergence in the Euclidean metric). \square

There is no such result as Theorem 3.8 for right-composition sequences. To see this, consider the function $g(z) = z + 1$ acting on the upper half-plane \mathbb{H} with Denjoy–Wolff point ∞ . (Here \mathbb{H} takes the place of the unit disc \mathbb{D} .) Let $h(z) = i + e^{2\pi iz}$, which is a holomorphic self-map of \mathbb{H} that satisfies $hg = h$. Now consider the right-composition sequence $G_n = g_1 g_2 \cdots g_n$, where $g_1 = h$ and $g_n = g$, for $n > 1$. Then (g_n) converges to g in the fastest possible way, but $G_n = hg^{n-1} = h$.

The following, similar example exhibits even worse behaviour of the sequence (G_n) . We provide only a sketch of the construction; the details will be filled in by Theorem 3.10 to follow. Our construction requires the theory of prime ends for more information on which we refer to [37, Section 17].

EXAMPLE 3.9. This example also uses \mathbb{H} rather than \mathbb{D} . We define $g(z) = z/2$, which is a holomorphic self-map of \mathbb{H} with Denjoy–Wolff point 0. Let D be the simply connected domain shown in Figure 3.1. It is obtained by removing two vertical line segments and various horizontal line segments from \mathbb{H} to leave an infinite snake-like domain, as shown in the figure. There are infinitely many horizontal line segments, and they accumulate at the real interval $[-1, 1]$, which is a prime end of D .

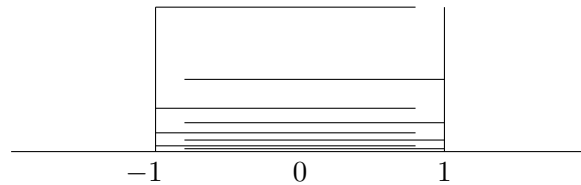


FIGURE 3.1. Domain D with a prime end at $[-1, 1]$

We define h to be a conformal map from \mathbb{H} to D . This map induces a one-to-one correspondence between the extended real line (the boundary of \mathbb{H}) and the prime ends of D . We choose h such that 0 corresponds to the prime end $[-1, 1]$. Now consider the right-composition sequence $G_n = g_1 g_2 \cdots g_n$, where $g_1 = h$ and $g_n = g$, for $n > 1$. Then (g_n) converges to g as quickly as possible, however, we will show that $(G_n(i))$ diverges. To see this, first observe that

$$G_n(i) = h g^n(i) = h(i/2^n), \quad \text{for } n = 1, 2, \dots$$

Since h is a conformal map from \mathbb{H} to D , it preserves hyperbolic distance between these two domains. So the hyperbolic length of the hyperbolic geodesic Γ_n between $G_{n-1}(i)$ and $G_n(i)$ in D is equal to the hyperbolic distance between $i/2^{n-1}$ and $i/2^n$ in \mathbb{H} , namely $\log 2$. Now, as n increases, $i/2^n$ approaches 0 (in the Euclidean metric), and $G_n(i)$ approaches the prime end $[-1, 1]$ (in the Euclidean metric). From the shape of D we can see that $(G_n(i))$ accumulates at an interval within $[-1, 1]$, so it diverges. \square

Example 3.9 indicates that there is little hope of obtaining a simple analogue of Theorem 3.8 for right-composition sequences. In fact, in the next theorem we show that the function g in Example 3.9 can be replaced with any holomorphic self-map of \mathbb{H} with Denjoy–Wolff point on the boundary. Theorem 3.10 does not feature in [18].

THEOREM 3.10. *Let g be a holomorphic self-map of \mathbb{D} with a Denjoy–Wolff point ζ on the boundary of \mathbb{D} . For every sequence of neighbourhoods $\mathcal{U}_1, \mathcal{U}_2, \dots$ of g in $\text{Hol}(\mathbb{D})$, there exists a sequence of holomorphic functions (g_n) such that $g_k \in \mathcal{U}_k$ for all $k = 1, 2, \dots$, and the right-composition sequence $G_n = g_1 g_2 \cdots g_n$ diverges.*

PROOF. The proof will be carried out in the upper half-plane, and without loss of generality we may assume that $\zeta = 0$. We are going to construct a function in $\text{Hol}(\mathbb{H})$ that has a prime end at 0, for which the impression is an interval on the real line.

Let (\mathcal{U}_n) be any sequence of neighbourhoods of g in $\text{Hol}(\mathbb{H})$. First, we construct a simply connected, proper subdomain D of \mathbb{H} , as shown in Figure 3.2.

Take two line segments $\gamma_1 = [-1, -1 + i]$ and $\gamma_2 = [1, 1 + i]$, orthogonal to the real line,

and for $k = 1, 2, \dots$ construct horizontal line segments δ_k , such that

$$\delta_k = \begin{cases} [-1 + \frac{i}{k}, 1 - \frac{1}{10} + \frac{i}{k}], & \text{for } k = 1, 3, 5, \dots \\ [-1 + \frac{1}{10} + \frac{i}{k}, 1 + \frac{i}{k}], & \text{for } k = 2, 4, 6, \dots \end{cases}$$

The desired simply connected domain is defined as

$$D = \mathbb{H} \setminus \left(\gamma_1 \cup \gamma_2 \cup \bigcup_{k=1}^{\infty} \delta_k \right).$$

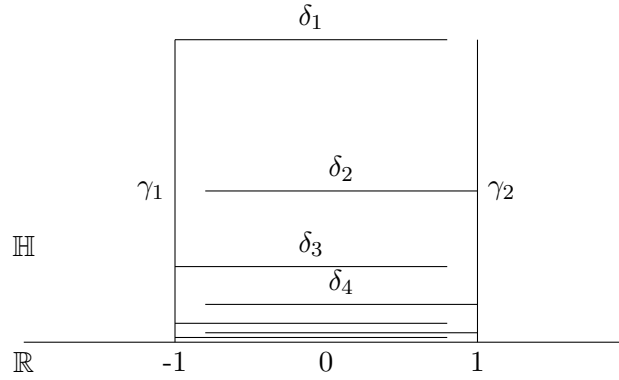


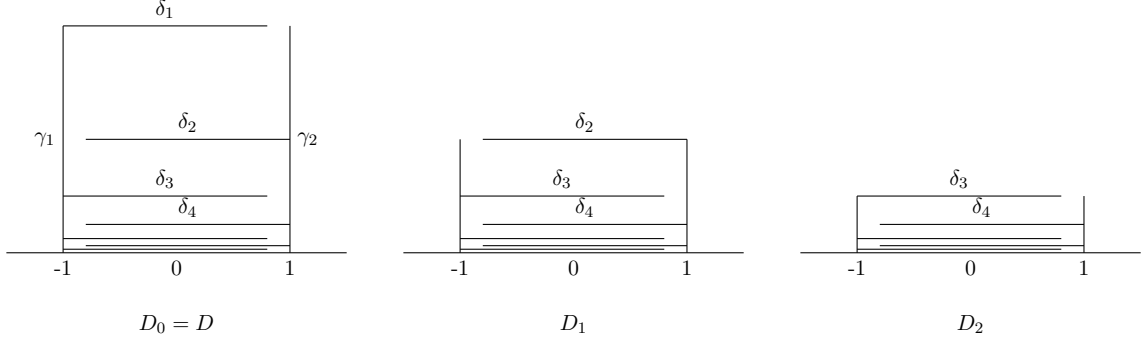
FIGURE 3.2. The domain D

Now, construct the following sequence of subdomains of \mathbb{H} (see Figure 3.3):

$$D_0 = D,$$

$$D_k = \mathbb{H} \setminus \left(\left\{ z \in \gamma_1 \cup \gamma_2 : \operatorname{Im} z \leq \frac{1}{k} \right\} \cup \bigcup_{i=k}^{\infty} \delta_i \right), \quad \text{for } k = 1, 2, \dots$$

All the domains D_n are simply connected and are missing a sequence of horizontal line segments that accumulate at an interval within $[-1, 1]$. We claim that the sequence of domains (D_n) converges to \mathbb{H} in terms of kernel convergence (for more details of kernel convergence and Carathéodory's kernel theorem, see [42, pages 13,14]). To prove this claim, observe that the point i lies inside D_k for all $k \geq 1$, and for any point $z \in \mathbb{H}$ and any neighbourhood U of z strictly contained in \mathbb{H} , the domains D_n contain U , for all n large enough. Furthermore, it is clear that if we choose any point w on the boundary of \mathbb{H} then for all $n = 1, 2, \dots$, there exists a sequence of points (a_n) in ∂D_n , such that a_n

FIGURE 3.3. The domains D_n

converges to w , in Euclidean terms, proving our claim.

Suppose that ϕ_n is the Riemann mapping from \mathbb{H} to D_n , with $\phi_n(i) = i$ and $\phi'_n(i) > 0$. Then we can use Carathéodory's kernel theorem to see that the sequence (ϕ_n) converges locally uniformly to the identity.

Suppose that $[K_1, E_1] \cap \dots \cap [K_m, E_m] \subset \mathcal{U}_1$ is an element in the base of the compact-open topology of $\text{Hol}(\mathbb{H})$. That is, the images $g(K_l)$ are contained in E_l , for every $l = 1, 2, \dots, m$, where K_l are compact and E_l open. Let K be a closed hyperbolic disc centred at i that contains K_l and $g(K_l)$, for all $l = 1, 2, \dots, m$. We can choose $\varepsilon > 0$ small enough, such that

$$\left\{ w \in \mathbb{H} : \sup_{z \in K_l} \rho_{\mathbb{H}}(w, g(z)) < \varepsilon \right\} \subset E_l, \quad \text{for every } l = 1, 2, \dots, m,$$

and by the convergence of (ϕ_n) to the identity, there exists a positive integer N so that

$$\sup_{z \in K} \rho_{\mathbb{H}}(\phi_N(z), z) < \frac{\varepsilon}{2}.$$

Recall that ϕ_N maps \mathbb{H} conformally onto the domain D_N . Thus, the function ϕ_N has a prime end at some point $p \in \partial\mathbb{H}$, whose impression is a nontrivial interval within $[-1, 1]$. So, if h is an elliptic Möbius transformation fixing \mathbb{H} , with $h(i) = i$ and $h(p) = 0$, then $\phi_N h^{-1}$ has a prime end at 0, whose impression is a nontrivial interval within $[-1, 1]$. Also, since $h(K) = K$, for $z \in K$ we have

$$\rho_{\mathbb{H}}(h\phi_N h^{-1}(z), z) = \rho_{\mathbb{H}}(\phi_N h^{-1}(z), h^{-1}(z)) < \frac{\varepsilon}{2},$$

which implies that for every $l = 1, 2, \dots, m$,

$$\sup_{z \in K_l} \rho_{\mathbb{H}}(h\phi_N h^{-1}g(z), g(z)) \leq \sup_{z \in K} \rho_{\mathbb{H}}(h\phi_N h^{-1}(z), z) < \varepsilon.$$

Hence, from our choice of ε , the images $h\phi_N h^{-1}g(K_l)$ are contained in E_l , for every $l = 1, 2, \dots, m$, and so the function $h\phi_N h^{-1}g$ lies in $[K_1, E_1] \cap \dots \cap [K_m, E_m] \subset \mathcal{U}_1$.

Now, define the sequence of functions

$$g_k = \begin{cases} h\phi_N h^{-1}g, & \text{for } k = 1 \\ g, & \text{otherwise,} \end{cases}$$

and note that $g_k \in \mathcal{U}_k$, for all k . Also, for any positive integer $n \geq 1$,

$$G_n = h\phi_N h^{-1}g^n.$$

We are now going to prove that the sequence $(h\phi_N h^{-1}g^n(i))$ diverges. Since h is a Möbius transformation, this is equivalent to proving that $z_n = \phi_N h^{-1}g^n(i)$ diverges, as $n \rightarrow \infty$.

Observe that the function $\phi_N h^{-1}g^n$ maps the domain D_N into itself, for all n , and the accumulations points of the sequence (z_n) lie in the interval $[-1, 1]$. Also, the hyperbolic distance, in the domain D_N , of two subsequent terms of the sequence (z_n) is bounded, as

$$(3.4.1) \quad \rho_{D_N}(z_n, z_{n+1}) = \rho_{D_N}(\phi_N h^{-1}g^n(i), \phi_N h^{-1}g^{n+1}(i)) \leq \rho_{D_N}(i, g(i)).$$

Let Γ_n be the hyperbolic geodesics in D_N joining the points z_n and z_{n+1} , for $n = 1, 2, \dots$. The domain D_N is a proper subdomain of \mathbb{H} , and so the hyperbolic metric $\lambda_{D_N}(z)|dz|$ is strictly greater than $\lambda_{\mathbb{H}}(z)|dz|$, for all $z \in D_N$. This implies that for $\ell_{D_N}(\Gamma_n)$ and $\ell_{\mathbb{H}}(\Gamma_n)$, the hyperbolic length of Γ_n in D_N and \mathbb{H} respectively, we have

$$(3.4.2) \quad \ell_{\mathbb{H}}(\Gamma_n) \leq \ell_{D_N}(\Gamma_n) = \rho_{D_N}(z_n, z_{n+1}), \quad \text{for all } n \in \mathbb{N}.$$

Recall that $\lambda_{\mathbb{H}}(z) = 1/\operatorname{Im} z$, and so (3.4.1) and (3.4.2) yield the following

$$(3.4.3) \quad \int_{\Gamma_n} \frac{1}{\operatorname{Im} z} |dz| = \ell_{\mathbb{H}}(\Gamma_n) \leq \rho_{D_N}(i, g(i)).$$

Also,

$$\sup_{z \in \Gamma_n} \{\operatorname{Im} z\} \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

Hence, (3.4.3) implies that the Euclidean length of the geodesics Γ_n tends to zero as $n \rightarrow \infty$. Therefore, using the geometry of the domain D_N it can easily be shown that the sequence z_n accumulates at an interval within $[-1, 1]$, as required. \square

CHAPTER 4

Universal constraints on semigroups of hyperbolic isometries

4.1. Semigroups of hyperbolic isometries

We now shift our focus to the isometries of the hyperbolic plane \mathbb{H} . As we saw in Chapter 1, the group of orientation-preserving hyperbolic isometries of \mathbb{H} is the group of real Möbius transformations $\mathrm{PSL}(2, \mathbb{R})$. Right-composition sequences of hyperbolic isometries appear naturally in the study of continued fractions, and, in fact, certain problems on the convergence of continued fractions can be restated as function theoretic problems for isometries (see, for example, [7]).

Motivated by this fact, Jacques and Short [26] studied the connections between the convergence of right-composition sequences generated by a finite collection of Möbius transformations \mathcal{F} , and the topological properties of the semigroup generated by \mathcal{F} in $\mathrm{PSL}(2, \mathbb{R})$. They proved ([26, Theorem 1.3]) that such a composition sequence converges to a constant on the boundary of \mathbb{H} , if and only if the semigroup generated by the collection \mathcal{F} does not accumulate to the identity function. Therefore, in this setting, the problem of identifying convergent composition sequences can be converted to a classification problem for semigroups of hyperbolic isometries.

Continuing the work of Jacques and Short, in this chapter we study the geometry of semigroups of conformal isometries of the hyperbolic plane. In particular, for a semigroup generated by a finite collection of isometries, we will calculate explicit geometric constraints on its generators which imply that elements of the semigroup do not accumulate to the identity function. The main result of this chapter is Theorem 4.6 to follow.

Semigroups of Möbius transformations have been previously studied by Fried, Marotta and Stankiewicz [22], as a particular branch of the theory of semigroups of rational functions initiated by Hinkkanen and Martin [25]. Our approach follows techniques similar to

those used by Jacques and Short [26], who further developed the material in [22], while incorporating well-known results from the theory of Fuchsian groups. Also, the theory of semigroups is closely related to the work of Avila, Bochi and Yoccoz [2] on uniformly hyperbolic finitely-valued $\mathrm{PSL}(2, \mathbb{R})$ -cocycles. In the last two chapters of this thesis we will further explore the connection between these two subjects.

Recall that the group $\mathrm{PSL}(2, \mathbb{R})$ consists of the Möbius transformations of the form $z \mapsto (az + b)/(cz + d)$, where $a, b, c, d \in \mathbb{R}$ and $ad - bc = 1$, and acts on the extended real line $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$. By the term *semigroup* we mean a subset of $\mathrm{PSL}(2, \mathbb{R})$ that is closed under composition. We say that a semigroup S is *generated* by a set $\mathcal{F} \subset \mathrm{PSL}(2, \mathbb{R})$, if every element of S can be written as a composition of transformations of \mathcal{F} . A semigroup is called *finitely-generated* if there exists a finite set of *generators* \mathcal{F} . If f_1, f_2, \dots, f_n are transformations in $\mathrm{PSL}(2, \mathbb{R})$, then $\langle f_1, f_2, \dots, f_n \rangle$ will denote the semigroup generated by $\{f_1, f_2, \dots, f_n\}$. A subset of $\mathrm{PSL}(2, \mathbb{R})$ is called *discrete* if the topology it inherits from $\mathrm{PSL}(2, \mathbb{R})$ is the discrete topology. For semigroups of Möbius transformations the authors of [26] also make the following definition.

DEFINITION 4.1. A semigroup S is called *semidiscrete* if the identity function is not an accumulation point of S . We say that S is *inverse-free* if it contains no inverses of elements of S .

For subgroups of $\mathrm{PSL}(2, \mathbb{R})$, the discrete and semidiscrete properties are equivalent (see, for example, [6, page 14] or [28, page 26]). This is not the case for semigroups, however, as the following example illustrates (see [26, Section 3]).

EXAMPLE 4.2 ([26]). Consider the transformations $f(z) = 2z$ and $g(z) = \frac{1}{2}z + 1$, and let S be the semigroup generated by f and g . It is easy to check that elements of $\langle f, g \rangle$, which are not iterates of f or g , are of the form $2^{n+m}z + 2^n + 2^{n-1}c$, for some $n \in \mathbb{N} \cup \{0\}, m \in \mathbb{Z}$ and $c \geq 0$. So, elements of S cannot accumulate at the identity and S is semidiscrete and inverse-free. It is not discrete, however, as

$$g^n f^n(z) = z + 2 - \frac{1}{2^{n-1}} \rightarrow z + 2, \quad \text{as } n \rightarrow \infty. \quad \square$$

For finitely-generated semigroups, the semidiscrete and inverse-free properties can be expressed geometrically, as we can see from the following theorem from [26, Theorem 7.1].

THEOREM 4.3 ([26]). *Suppose that S is a semigroup generated by a finite collection \mathcal{F} of Möbius transformations. Then S is semidiscrete and inverse-free if and only if there is a nontrivial closed subset X of $\overline{\mathbb{H}}$ that is mapped strictly inside itself by each member of \mathcal{F} .*

Using Theorem 4.3, the semidiscrete and inverse-free properties of Example 4.2 can be easily deduced by the fact that the semigroup $\langle f, g \rangle$ in Example 4.2 maps the interval $[1, \infty]$ strictly inside itself. The result ([26, Theorem 1.3]) that connects the behaviour of composition sequences to the topological properties of the semigroup can be stated as follows.

THEOREM 4.4 ([26]). *Suppose that S is a finitely generated semigroup. Then, S is semidiscrete and inverse-free if and only if every right-composition sequence in S converges to a point in $\overline{\mathbb{R}}$.*

We now introduce a class of finitely generated semigroups that will prove important to our analysis on this, as well as the two following chapters.

DEFINITION 4.5. Let S be a semigroup generated by a finite collection of Möbius transformations \mathcal{F} . If there exists a union C of m open intervals in $\overline{\mathbb{R}}$, with disjoint closures, such that each element of \mathcal{F} maps C strictly inside itself, then S is called a *Schottky semigroup*.

Suppose that the integer m in Definition 4.5 is the smallest integer with this property, in the sense that if there exists another union D of n open intervals in $\overline{\mathbb{R}}$, with disjoint closures, such that each element of \mathcal{F} maps D strictly inside itself, then $m \leq n$. We then say that S is a Schottky semigroup of rank m .

By Theorem 4.3, every Schottky semigroup is semidiscrete and inverse-free. Also, observe that Example 4.2 is a Schottky semigroup of rank one. The terminology of this definition comes from the connection of Schottky semigroups to a large class of discrete groups called Schottky groups, which we will introduce in Section 6.3. Note however that not every Schottky semigroup is contained in a discrete group, as Example 4.2 indicates.

For two hyperbolic transformations f, g , we define the *cross ratio* $C(f, g)$ of f and g to be

$$C(f, g) = \frac{\alpha(f) - \alpha(g)}{\alpha(f) - \beta(g)} \frac{\beta(f) - \beta(g)}{\beta(f) - \alpha(g)}.$$

Since the fixed points of f and g lie in the same extended Euclidean line in $\overline{\mathbb{C}}$, we have that $C(f, g)$ lies in $\overline{\mathbb{R}}$. The value of $C(f, g)$ allows us to infer information about the geometric configuration of $\text{Ax}(f)$ and $\text{Ax}(g)$, as it is shown in Figure 4.1 (the configurations shown in this figure are justified by Lemma 4.15 to follow). It is easy to check that for two hyperbolic transformations f, g , the cross ratio $C(f, g)$ is never one. Also, $C(f, g) = 0$ if and only if either $\alpha(f) = \alpha(g)$ or $\beta(f) = \beta(g)$. Also, $C(f, g) = \infty$ if and only if $\alpha(f) = \beta(g)$ or $\beta(f) = \alpha(g)$.

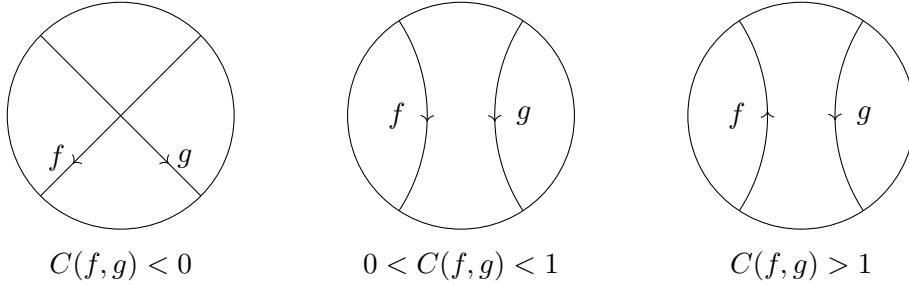


FIGURE 4.1. Cross ratio and geometric configuration

The main goal of this chapter is to prove the following result.

THEOREM 4.6. *Let \mathcal{F} be a finite collection of hyperbolic transformations such that all the fixed points of transformations in \mathcal{F} are distinct, and $S = \langle \mathcal{F} \rangle$ is not a Schottky semigroup of rank one.*

(1) *Suppose that*

$$\tau(h) < \frac{1}{5} \min \left\{ \frac{C(f, g) - 1}{C(f, g) + 3}, 1 \right\}, \quad \text{for all } h \in \mathcal{F},$$

where the minimum is taken over all pairs f, g in \mathcal{F} with $C(f, g) > 1$. Then S is not semidiscrete.

(2) *Suppose that*

$$\tau(h) > 4 \max \left\{ |\log |C(f, g)|| + \log \frac{\sqrt{|C(f, g)|} + 1}{|\sqrt{|C(f, g)|} - 1|} \right\} + 3, \quad \text{for all } h \in \mathcal{F},$$

where the maximum is taken over all pairs f, g in \mathcal{F} with $C(f, g) \neq -1$. Then S is a Schottky semigroup.

Observe that if the fixed points of the hyperbolic transformations f, g are distinct, then $C(f, g) \notin \{0, 1, \infty\}$ and so the constants in Theorem 4.6 are well-defined.

Given a collection of hyperbolic transformations $\{f_1, f_2, \dots, f_n\}$, with distinct fixed points, Theorem 4.6 provides us with constraints on the translation length of each f_i in order for the semigroup $\langle f_1, f_2, \dots, f_n \rangle$ to be semidiscrete and inverse-free. A strength of Theorem 4.6 is that these constraints are given by simple algebraic formulas which can be computed by considering the generators f_1, f_2, \dots, f_n in pairs.

Our main result also holds for some semigroups with generators whose fixed points meet, but the proof we have so far does not produce a simple constant and requires further investigation. This is work in progress.

We now present an example where we compute the constants in Theorem 4.6 explicitly. Suppose that the axes of the hyperbolic transformations f_1, f_2, \dots, f_5 are as shown in Figure 4.2 and consider the collection $\mathcal{F} = \{f_1, f_2, \dots, f_5\}$.

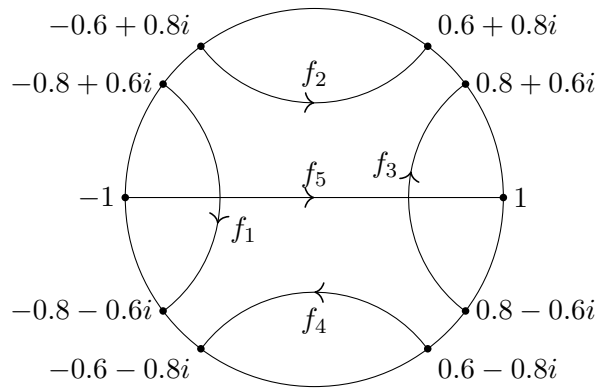


FIGURE 4.2. An application of Theorem 4.6

The cross ratios of all the pairs are:

$$\begin{aligned} C(f_1, f_2) &= C(f_2, f_3) = C(f_3, f_4) = C(f_4, f_1) = \frac{7}{25}, \\ C(f_1, f_3) &= C(f_2, f_4) = \frac{25}{4}, \\ C(f_1, f_5) &= C(f_3, f_5) = -1, \\ C(f_4, f_5) &= \frac{1}{C(f_2, f_5)} = 5. \end{aligned}$$

Therefore, if

$$\tau(f_k) < \frac{1}{5} \min \left\{ \frac{5-1}{5+3}, 1 \right\} = \frac{1}{5 \cdot 2} = 0.1, \text{ for } k = 1, 2, \dots, 5,$$

then the semigroup $\langle \mathcal{F} \rangle$ is not semidiscrete, whereas if

$$\tau(f_k) > 4 \left(\log \frac{25}{4} + \log \frac{\sqrt{\frac{25}{4} + 1}}{\sqrt{\frac{25}{4} - 1}} \right) + 3 = 4 \log \frac{175}{12} + 3 \approx 13.719, \text{ for } k = 1, 2, \dots, 5,$$

then $\langle \mathcal{F} \rangle$ is semidiscrete and inverse-free, and in fact it is a Schottky semigroup of rank two.

4.2. Examples and preliminary results

In order to further motivate the results in this chapter, let us describe certain examples of semigroups that exhibit interesting behaviour, but do not fit the framework of Theorem 4.6.

In accordance with the theory of Fuchsian groups, we make the following definition.

DEFINITION 4.7. A semigroup S is called *elementary* if it has a finite orbit in $\overline{\mathbb{H}}$.

One can easily verify that S is elementary if and only if either all its elements have a common fixed point in $\overline{\mathbb{H}}$, or there is a pair of points that is fixed as a set by all elements of S . If S is not elementary, we will say that it is *non-elementary*.

The next lemma, taken from [26, Lemma 10.4], describes a class of elementary, two-generator semigroups which illustrates that a semidiscrete semigroup can fail to be discrete in a very strong sense. Note that a hyperbolic transformation $az+b$, with $a \neq 1$ and $b \in \mathbb{R}$, fixes the finite point $b/(1-a)$, and this point is attracting if $a < 1$ and repelling if $a > 1$.

LEMMA 4.8 ([26]). *Suppose $f(z) = az + b$ and $g(z) = cz + d$ are two hyperbolic transformations, with $1 < a$, $0 < c < 1$ and $b/(1 - a) < d/(1 - c)$. The semigroup S generated by f and g is semidiscrete, and the closure of S in $\mathrm{PSL}(2, \mathbb{R})$ contains all transformations of the form $z + t$, for $t \geq d/(1 - c) - b/(1 - a)$.*

Observe that the example mentioned in Section 4.1 now comes as a special case of Lemma 4.8. Elementary semigroups were studied in [26, Section 10], where it was shown that all elementary semigroups generated by a finite collection of hyperbolic transformations are semidiscrete and inverse-free, apart from those that contain either two hyperbolic transformations with the same axis, or the three hyperbolic transformations described in the following lemma found in [26, Corollary 10.5].

LEMMA 4.9 ([26]). *Let S be the semigroup generated by $f_i = a_i z + b_i$, for $i = 1, 2, 3$, with $a_1, a_3 > 1$ and $0 < a_2 < 1$, and*

$$\frac{b_1}{1 - a_1} \leq \frac{b_2}{1 - a_2} \leq \frac{b_3}{1 - a_3}.$$

Then S is not semidiscrete and inverse-free.

Before we move on, let us introduce the notion of limit sets for Möbius semigroups. We define the *forward limit set* $\Lambda^+(S)$ of a semigroup S to be the set accumulation points of $\{f(z_0) : f \in S\}$ in $\overline{\mathbb{H}}$, where $z_0 \in \mathbb{H}$, with respect to the chordal metric in $\overline{\mathbb{C}}$. The *backward limit set* $\Lambda^-(S)$ of S is defined to be the forward limit set of $S^{-1} = \{f^{-1} : f \in S\}$. It is easy to check that the limit sets are independent of the choice of z_0 , and $\Lambda^+(S)$ is forward invariant under transformations in S whereas $\Lambda^-(S)$ is backward invariant. Furthermore, Fried, Marotta and Stankewitz [22, Theorem 2.4, Proposition 2.6, and Remark 2.20] showed that $\Lambda^-(S)$ is the closure of all the repelling fixed points of hyperbolic elements of S . Equivalently, $\Lambda^+(S)$ is the closure of all the attracting fixed points of hyperbolic transformations in S .

As an example, we will evaluate the limit sets of the two-generator semigroup introduced in Lemma 4.8.

LEMMA 4.10. *Let S be the semigroup generated by $f(z) = az$ and $g(z) = cz + d$, where $0 < c < 1 < a$ and $0 < d/(1 - c)$. The forward limit set of S is the closed interval $[d/(1 - c), \infty]$.*

PROOF. Note that the interval $[d/(1 - c), \infty]$ is mapped into itself by f and g and so it has to contain all the attracting fixed points of hyperbolic elements of S . Thus $\Lambda^+(S) \subseteq [d/(1 - c), \infty]$. Also, observe that

$$(4.2.1) \quad g^m f^n = a^n c^m z + \frac{d}{1 - c}(1 - c^m),$$

for any $m, n \in \mathbb{N}$. Suppose, first, that $a^n c^m \neq 1$, for all positive integers m, n and let $0 < \lambda < 1$. Then for every $\varepsilon > 0$, there exist m_0 and n_0 such that $|a^{n_0} c^{m_0} - \lambda| < \varepsilon$ and $|\frac{d}{1 - c} c^{m_0}| < \varepsilon$. Therefore,

$$\left| g^{m_0} f^{n_0}(z) - \left(\lambda z + \frac{d}{1 - c} \right) \right| < \varepsilon(|z| + 1),$$

for all $z \in \mathbb{H}$. Observe that the finite fixed point of the transformation $\lambda z + d/(1 - c)$ is $\frac{d}{(1 - c)(1 - \lambda)}$, and it is an attracting fixed point because $\lambda < 1$. We deduce that in this case, $\frac{d}{(1 - c)(1 - \lambda)}$ lies in $\Lambda^+(S)$, for all $0 < \lambda < 1$, which yields the desired result.

If on the other hand, $a^\mu b^\nu = 1$ for some $\mu, \nu \in \mathbb{N}$ then, by considering the subsemigroup $\langle f^\nu, g^\mu \rangle$ of S , we can assume that $ac = 1$. So, from (4.2.1)

$$H_n(z) = g^n f^n(z) = z + \frac{d}{1 - c}(1 - c^n),$$

for all $n \in \mathbb{N}$. So for all non-negative integers n, l , the transformation

$$g^{2n} H_n^l f^n(z) = c^n z + l \frac{d}{1 - c}(1 - c^n) c^{2n} + \frac{d}{1 - c}(1 - c^{2n}),$$

is a hyperbolic transformation in S , whose attracting fixed point is

$$\frac{d}{1 - c}(lc^{2n} + c^n + 1).$$

For every real number $x > 0$, and every $\varepsilon > 0$, we can choose positive integers l, n , so that $|lc^{2n} - x| < \frac{\varepsilon}{2}$ and $c^n < \frac{\varepsilon}{2}$. This implies that for every $r > 1$, and all $\varepsilon > 0$, we can find l, n such that $|lc^{2n} + c^n + 1 - r| < \varepsilon$. Hence, the point $rd/(1 - c)$ lies in $\Lambda^+(S)$ for all $r > 1$. \square

Observe that a semigroup S is inverse-free if and only if $S \cap S^{-1} = \emptyset$. It is also interesting to note that $S \cap S^{-1}$, if non-empty, is a group whose limit set lies in the intersection $\Lambda^+(S) \cap \Lambda^-(S)$. So, one might expect that there is a connection between the size of the intersection of the limit sets and the size of $S \cap S^{-1}$. Obviously, if S is a group then the backward and forward limit sets coincide. For finitely-generated, non-elementary semigroups, the following theorem [26, Theorem 1.9] shows that the other direction of this statement is also true.

THEOREM 4.11 ([26]). *Let S be a finitely-generated, non-elementary and semidiscrete semigroup. If $\Lambda^-(S) \subseteq \Lambda^+(S)$, then S is a group.*

Using Theorem 4.11 we establish the the following lemma, where $\Lambda^+(S)^\circ$ denotes the interior of the forward limit set.

LEMMA 4.12. *Let S be a finitely-generated, non-elementary semigroup that is not a discrete group, and suppose that $\Lambda^+(S)^\circ \cap \Lambda^-(S) \neq \emptyset$. Then S is not semidiscrete.*

PROOF. Since $\Lambda^-(S)$ is the smallest closed set that contains all the repelling fixed points of hyperbolic elements of S , there exists a hyperbolic transformation f in S with repelling fixed point in $\Lambda^+(S)^\circ$. So, by the invariance of the limit sets under the semigroup, we have that $\overline{\mathbb{R}} = \overline{\{f^n(\Lambda^+(S)^\circ) : n \in \mathbb{N}\}} \subseteq \Lambda^+(S)$ and therefore $\Lambda^+(S) = \overline{\mathbb{R}}$. Finally, because S is non-elementary and $\Lambda^-(S) \subset \overline{\mathbb{R}} = \Lambda^+(S)$, Theorem 4.11 tells us that if S were semidiscrete, it would have to be a discrete group, which is a contradiction. \square

We end this section by introducing a class of three-generator semigroups that contains the semigroup described in Lemma 4.9.

LEMMA 4.13. *Let $f(z) = az$, $g(z) = cz + d$ and h be hyperbolic transformations, where $0 < c < 1 < a$ and $0 < d/(1 - c)$.*

- (1) *If $\beta(h) \in (d/(1 - c), \infty)$, then $\langle f, g, h \rangle$ is not semidiscrete.*
- (2) *If $\beta(h) = d/(1 - c)$, then $\langle f, g, h \rangle$ is semidiscrete if and only if $\alpha(h)$ lies in the open interval $(d/(1 - c), \infty)$.*

PROOF. First, observe that Lemma 4.10 implies $\Lambda^+(\langle f, g \rangle) = [d/(1 - c), \infty]$. Assume that $\beta(h) \in (d/(1 - c), \infty) = \Lambda^+(\langle f, g \rangle)^\circ$. Then because $\beta(h)$ lies in $\Lambda^+(\langle f, g \rangle)^\circ \subseteq$

$\Lambda^+(\langle f, g, h \rangle)^\circ$, the semigroup generated by f, g and h satisfies the assumptions of Lemma 4.12, and thus it is not semidiscrete.

Suppose that $\beta(h) = d/(1 - c)$. If the attracting fixed point $\alpha(h)$ lies in $[\infty, d/(1 - c)]$, then we can apply a modification of Lemma 4.10 in order to deduce that $\Lambda^+(\langle g, h \rangle) = [\alpha(h), d/(1 - c)]$. Hence $[\alpha(h), \infty] \subseteq \Lambda^+(\langle f, g, h \rangle)$. Thus, $\langle f, g, h \rangle$ is not semidiscrete since $\beta(h)$ lies in the interior of the forward limit set of $\langle f, g, h \rangle$. Assume that $\alpha(h) \in (d/(1 - c), \infty)$. Define $K = \{z \in \overline{\mathbb{H}} : \operatorname{Re} z \geq d/(1 - c) \text{ and } \operatorname{Re} z \leq \operatorname{Im} z\}$. It is easy to check that K is a closed subset of $\overline{\mathbb{H}}$ that is mapped strictly inside itself by f, g and h , and Theorem 4.3 yields the desired result. \square

See Figure 4.3 for examples of axis configurations that satisfy the hypotheses of Corollary 4.13. The first three semigroups fall under the first case of Corollary 4.13 and are not semidiscrete, whereas the one on the far right is semidiscrete.

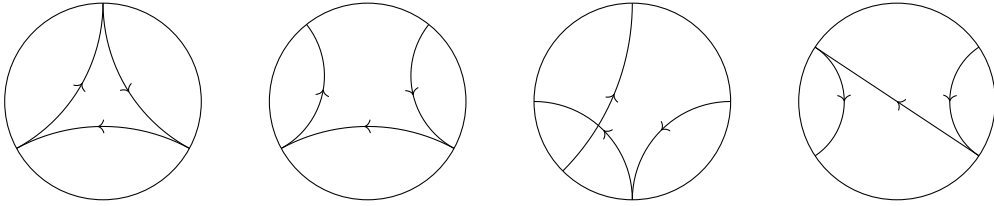


FIGURE 4.3. Examples of semigroups in Lemma 4.13.

The next result is a corollary of Lemma 4.13.

COROLLARY 4.14. *Let $f(z) = az$, $g(z) = cz + d$ and h be hyperbolic transformations, where $0 < c < 1 < a$ and $0 < d/(1 - c)$.*

- (1) *If $\alpha(h) \in (\infty, 0)$, then $\langle f, g, h \rangle$ is not semidiscrete.*
- (2) *If $\alpha(h) = 0$, then $\langle f, g, h \rangle$ is semidiscrete if and only if $\beta(h)$ lies in the open interval $(\infty, 0)$.*

PROOF. Consider the semigroup $\langle f^{-1}, g^{-1}, h^{-1} \rangle$. Observe that $\langle f^{-1}, g^{-1}, h^{-1} \rangle$ is semidiscrete if and only if the semigroup $\langle f, g, h \rangle$ is semidiscrete. Hence, it suffices to prove the desired results for the semigroup $\langle f^{-1}, g^{-1}, h^{-1} \rangle$.

Suppose first that $\alpha(h) \in (\infty, 0) = (\beta(g), \beta(f))$. This is equivalent to $\beta(h^{-1}) \in (\alpha(g^{-1}), \alpha(f^{-1}))$,

and so we can apply a modification of the first part of Lemma 4.13 to the semigroup $\langle f^{-1}, g^{-1}, h^{-1} \rangle$ in order to deduce that it is not semidiscrete.

The second part of the corollary follows from similar arguments. \square

4.3. Two-generator semigroups

In this section we prove a version of Theorem 4.6 for two hyperbolic transformations f and g . Semigroups generated by pairs of Möbius transformations were studied by Avila, Bochi and Yoccoz [2, Chapter 3] and Jacques and Short [26, Section 12]. Our results will be obtained by following techniques similar to [26, Theorem 1.4] and modifying well-known results for two-generator Fuchsian groups (see [6, Chapter 11]).

Recall that for a transformation $h(z) = (az + b)/(cz + d)$ with $ad - bc = 1$, the trace of h is the number $\text{tr}(h) = a + d$. We will say that an interval $I \subset \overline{\mathbb{R}}$ is *symmetric with respect to* a hyperbolic transformation h , if the hyperbolic geodesic with the same endpoints as I , is perpendicular to the axis of h . Finally, if the axes of two transformations f and g cross at a point $p \in \mathbb{H}$, we define the angle $\theta \in [0, \pi]$ between $\text{Ax}(f)$ and $\text{Ax}(g)$ to be the angle at p of the hyperbolic triangle defined by $\alpha(f), \alpha(g)$ and p . In this case, we say that the axes of f and g *cross at an angle* θ .

We start by establishing formulas that relate the cross ratio $C(f, g)$ of two hyperbolic transformations f and g with the geometric configuration of their axes. Throughout, two hyperbolic lines will be called *disjoint* if they do not cross in \mathbb{H} and have distinct endpoints.

LEMMA 4.15. *Suppose that f and g are hyperbolic transformations.*

- (1) *If the axes of f and g cross at an angle θ , then $C(f, g) = -\tan^2 \frac{1}{2}\theta$.*
- (2) *If the axes of f and g are disjoint and a hyperbolic distance d apart, then*

$$C(f, g) = \begin{cases} \tanh^2 \frac{1}{2}d, & \text{if } 0 < C(f, g) < 1, \\ \coth^2 \frac{1}{2}d, & \text{if } C(f, g) > 1. \end{cases}$$

PROOF. First, observe that the cross ratio of four points in \mathbb{C} is invariant under conjugation by a Möbius transformation in $\text{PSL}(2, \mathbb{C})$.

For part (i), suppose that the axes of f and g cross at an angle θ . By conjugating f and g with a Möbius transformation in $\text{PSL}(2, \mathbb{R})$ so that $f(z) = az$ for some $a \neq 1$,

we can see that $C(f, g)$ has to be negative. Now, conjugate f and g by a transformation $\phi \in \text{PSL}(2, \mathbb{C})$ so that they act on the unit disc, and their axes meet at the origin (see Figure 4.4 on the left). Denote $F = \phi \circ f \circ \phi^{-1}$, $G = \phi \circ g \circ \phi^{-1}$, and observe that $C(F, G) = C(f, g)$. Because $|C(f, g)| = -C(f, g)$, it is easy to see that

$$C(F, G) = -\frac{|\alpha(F) - \alpha(G)|^2}{|\beta(G) - \alpha(F)|^2}.$$

Also, by the law of cosines, $|\alpha(F) - \alpha(G)|^2 = 2(1 - \cos \theta)$ and $|\beta(G) - \alpha(F)|^2 = 2(1 - \cos(\pi - \theta))$, which yield the desired equation.

Assume, now, that the axes of f and g are disjoint and a hyperbolic distance d apart. Note that in this case we have $C(f^{-1}, g) = 1/C(f, g)$, and so it suffices to assume that $C(f, g) > 1$. Conjugate f and g so that f fixes $-\sigma, \sigma$ and g fixes $-\lambda, \lambda$, for some $0 < \sigma < \lambda$ (see Figure 4.4 on the right). Then,

$$C(f, g) = \frac{(\lambda - \sigma)^2}{(\lambda + \sigma)^2}.$$

Observe that $d = \log \frac{\lambda}{\sigma}$, which implies that

$$\cosh d = \frac{\lambda^2 + \sigma^2}{2\lambda\sigma},$$

and using the half-angle formula for the hyperbolic cotangent completes our proof. \square

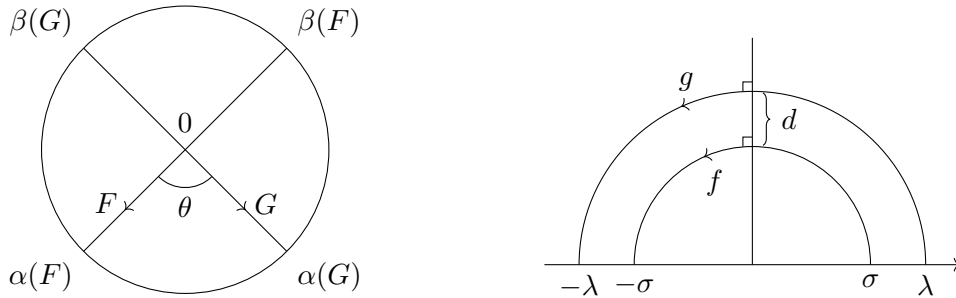


FIGURE 4.4. Formulas for the cross ratio of two hyperbolic transformations

We first consider the case of two hyperbolic transformations with disjoint axes. In order to prove Theorem 4.18, to follow, we will make use of the next theorem that can be directly inferred from Theorem 1.4, Theorem 1.5 and Lemma 12.9 from [26].

THEOREM 4.16 ([26]). *Suppose that f and g are hyperbolic transformations. Then the semigroup $\langle f, g \rangle$ satisfies one of the following possibilities: if $\langle f, g \rangle$ does not contain elliptic elements then it is semidiscrete and inverse-free; otherwise, either $\langle f, g \rangle$ is itself a discrete group, or else it is not semidiscrete.*

LEMMA 4.17. *Let f, g be two hyperbolic transformations with $C(f, g) > 1$, whose axes are a hyperbolic distance d apart. Suppose that*

$$\sinh \frac{1}{2}\tau(f), \sinh \frac{1}{2}\tau(g) > \frac{1}{\sinh \frac{1}{2}d}.$$

Then there exist open intervals A_f, B_f, A_g, B_g in $\overline{\mathbb{R}}$, with pairwise disjoint closures, that satisfy the following properties: the intervals A_f, B_f are symmetric with respect to f and $f(B_f^c)$ is contained in A_f ; the intervals A_g, B_g are symmetric with respect to g and $g(B_g^c)$ is contained in A_g . In particular $\langle f, g \rangle$ is a Schottky semigroup of rank two.

PROOF. Suppose that f and g are hyperbolic transformations with $C(f, g) > 1$, and let d be the hyperbolic distance between their axes. Let ℓ be the unique hyperbolic line that is perpendicular to the axes of f and g , and σ the reflection in ℓ . Also, define the reflections $\sigma_f = f \circ \sigma$ and $\sigma_g = \sigma \circ g$ and let ℓ_f, ℓ_g be their lines of reflection respectively (see Figure 4.5).

Consider the function $h: \mathbb{R}_+^2 \rightarrow \mathbb{R}$, with $h(x, y) = \cosh d \sinh x \sinh y - \cosh x \cosh y$, and suppose first that $\tau(f) = \tau(g) = x_0$, with

$$\sinh \frac{1}{2}x_0 = \frac{1}{\sinh \frac{1}{2}d}.$$

Then $h(\frac{1}{2}x_0, \frac{1}{2}x_0) = 1$ and it is easy to check that $h(x, \frac{1}{2}x_0)$ and $h(x, x)$ are increasing functions of x . Therefore, if $x > x_0$ then $h(x, \frac{1}{2}x_0), h(x, x) > 1$. By the symmetry of h we have that all points $(x, y) \in \mathbb{R}^2$ with $x, y > x_0$ satisfy $h(x, y) > 1$. Or equivalently if

$$(4.3.1) \quad \sinh \frac{1}{2}\tau(f), \sinh \frac{1}{2}\tau(g) > \frac{1}{\sinh \frac{1}{2}d},$$

we have that

$$(4.3.2) \quad \cosh d \sinh \frac{1}{2}\tau(f) \sinh \frac{1}{2}\tau(g) - \cosh \frac{1}{2}\tau(f) \cosh \frac{1}{2}\tau(g) > 1.$$

Suppose now that inequality (4.3.1) is satisfied. We are going to show that the desired intervals can be chosen as follows. Let A_f be the open interval in $\overline{\mathbb{R}}$ with the same endpoints as ℓ_f and containing α_f , and B_g the open interval with the same endpoints as ℓ_g and containing β_g . Also, let $B_f = \sigma(A_f)$ and $A_g = \sigma(B_g)$. It is obvious that $f(B_f^c) \subset A_f$ and $g(B_g^c) \subset A_g$, and so it suffices to prove that A_f and B_g have disjoint closures, as that would imply that the same holds for B_f and A_g .

If the line ℓ_f meets $\text{Ax}(g)$ in $\overline{\mathbb{H}}$, then the lines $\ell_f, \text{Ax}(f), \ell$ and $\text{Ax}(g)$ form a quadrilateral in \mathbb{H} , and so [6, Theorem 7.17.1] implies that

$$\sinh \frac{1}{2}\tau(f) \leq \frac{1}{\sinh d},$$

which is a contradiction since $\sinh \frac{1}{2}d < \sinh d$. Therefore, ℓ_f does not cross $\text{Ax}(g)$, and similarly ℓ_g does not cross $\text{Ax}(f)$.

So, we can now see that if $\overline{A_f} \cap \overline{B_g} \neq \emptyset$, then the lines $\ell, \text{Ax}(f), \ell_f, \ell_g$ and $\text{Ax}(g)$ define a pentagon in $\overline{\mathbb{H}}$, and thus [6, Theorem 7.18.1] implies that

$$\cosh d \sinh \frac{1}{2}\tau(f) \sinh \frac{1}{2}\tau(g) - \cosh \frac{1}{2}\tau(f) \cosh \frac{1}{2}\tau(g) \leq 1.$$

This, however, contradicts (4.3.2) and therefore $\overline{A_f} \cap \overline{B_g} = \emptyset$. □

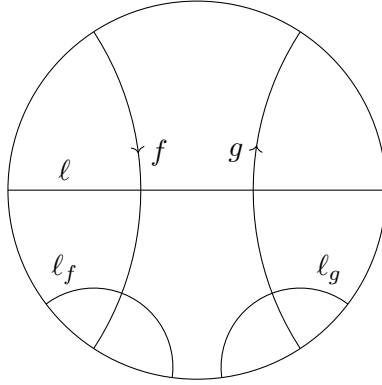


FIGURE 4.5. Hyperbolic transformations as products of reflections.

$$h: \mathbb{R}_+^2 \rightarrow \mathbb{R}, \text{ with}$$

THEOREM 4.18. *Let f, g be two hyperbolic transformations with $C(f, g) > 1$.*

(1) Suppose that

$$\tau(f), \tau(g) < \frac{1}{5} \frac{C(f, g) - 1}{C(f, g) + 3}.$$

Then $\langle f, g \rangle$ is not semidiscrete.

(2) Suppose that

$$\tau(f), \tau(g) > \log C(f, g) + \frac{3}{2}.$$

Then $\langle f, g \rangle$ is a Schottky semigroup of rank two.

PROOF. Let d be the hyperbolic distance between the axes of f and g . The proof revolves around evaluating the trace of the composition of f and g , which is given by the following equation found in [6, Theorem 7.38.3]:

$$\frac{1}{2} |\operatorname{tr}(f \circ g)| = \left| \cosh d \sinh \frac{1}{2} \tau(f) \sinh \frac{1}{2} \tau(g) - \cosh \frac{1}{2} \tau(f) \cosh \frac{1}{2} \tau(g) \right|.$$

Recall that $f \circ g$ is elliptic if $|\operatorname{tr}(f \circ g)| < 2$ and hyperbolic if $|\operatorname{tr}(f \circ g)| > 2$. Let \mathbb{R}_+^2 be the first quadrant of \mathbb{R}^2 and consider the function $h(x, y) = \cosh d \sinh x \sinh y - \cosh x \cosh y$ that was introduced in Lemma 4.17. Suppose that

$$(4.3.3) \quad \tau(f), \tau(g) < \frac{1}{5} \frac{C(f, g) - 1}{C(f, g) + 3}.$$

We are going to prove that there exist positive integers m, n , such that $f^m \circ g^n$ is elliptic, and $\langle f, g \rangle$ is not a discrete group. Then, Theorem 4.16 would imply that $\langle f, g \rangle$ is not semidiscrete.

First, we refer to [6, Theorem 11.6.9] which states that if $h(\frac{1}{2}\tau(f), \frac{1}{2}\tau(g)) < -\frac{1}{2}$, then the group generated by f and g is not discrete. Therefore, it suffices to prove that if f and g satisfy (4.3.3), then there exist positive integers m, n , such that $-1 < h(\frac{m}{2}\tau(f), \frac{n}{2}\tau(g)) < -\frac{1}{2}$.

Define $D = \{(x, y) \in \mathbb{R}_+^2 : -1 < h(x, y) < -\frac{1}{2}\}$, and let $b \geq 0$ be such that $h(b, b) = -\frac{1}{2}$.

Solving this equation for $\sinh b$ yields

$$\sinh b = \frac{1}{\sqrt{2}\sqrt{\cosh d - 1}} = \frac{1}{2 \sinh \frac{1}{2}d}.$$

Also, define $a \geq 0$ to be the unique solution of $h(a, a) = -\frac{7}{9}$. Then

$$\sinh a = \frac{\sqrt{2}}{3\sqrt{\cosh d - 1}} = \frac{1}{3 \sinh \frac{1}{2}d},$$

which implies that $a < b$. Let C be the square with vertices $(a, b), (b, b), (b, a), (a, a)$ (see Figure 4.6). We will prove that the compact set K bounded by the square C lies in \overline{D} . Note that points (x, y) in K satisfy the inequalities $\sinh a \leq \sinh x, \sinh y \leq \sinh b$, or equivalently $\cosh a \leq \cosh x, \cosh y \leq \cosh b$, where

$$\cosh a = \frac{\sqrt{9 \cosh d - 7}}{3\sqrt{\cosh d - 1}} \quad \text{and} \quad \cosh b = \frac{\sqrt{2 \cosh d - 1}}{\sqrt{2}\sqrt{\cosh d - 1}}.$$

We are going to show that h is increasing on vertical and horizontal line segments in K ,

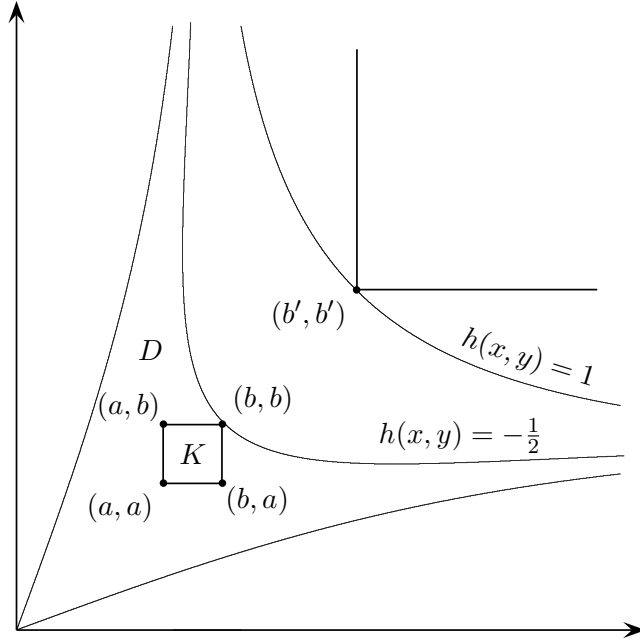


FIGURE 4.6. The square K inside the domain D , and the quarter-plane Q

which will complete the proof of our claim. Because h is symmetric with respect to the line $y = x$,

$$\frac{\partial h}{\partial x}(x, y) = \frac{\partial h}{\partial y}(y, x) = \cosh d \cosh x \sinh y - \sinh x \cosh y.$$

Thus, for all $(x, y) \in K$ we have that $\frac{\partial h}{\partial x}(x, y) \geq \cosh d \cosh a \sinh a - \sinh b \cosh b$, which is

$$\frac{\partial h}{\partial x}(x, y) \geq \cosh d \frac{\sqrt{9 \cosh d - 7}}{3\sqrt{\cosh d - 1}} \frac{\sqrt{2}}{3\sqrt{\cosh d - 1}} - \frac{1}{\sqrt{2}\sqrt{\cosh d - 1}} \frac{\sqrt{2}\sqrt{2 \cosh d - 1}}{2\sqrt{\cosh d - 1}}.$$

So, we have that

$$\frac{\partial h}{\partial x}(x, y) \geq \sqrt{2} \frac{\cosh d \sqrt{9 \cosh d - 7} - \sqrt{4 \cosh d - 2}}{9(\cosh d - 1)}.$$

Since $\sqrt{9 \cosh d - 7} > \sqrt{4 \cosh d - 2}$, the function h is increasing on horizontal and vertical line segments inside K .

Therefore, if $\tau(f), \tau(g) < 2(b - a)$ then there exist positive integers m, n , such that the point $(\frac{m}{2}\tau(f), \frac{n}{2}\tau(g))$ lies in the interior of K . In order to finish the proof of the first part we are going to show that

$$\frac{1}{5} \frac{C(f, g) - 1}{C(f, g) + 3} < 2(b - a).$$

Observe that

$$\begin{aligned} \sinh(b - a) &= \sinh b \cosh a - \cosh b \sinh a = \frac{\sinh^2 b \cosh^2 a - \cosh^2 b \sinh^2 a}{\sinh b \cosh a + \cosh b \sinh a} \\ &> \frac{\sinh^2 b \cosh^2 a - \cosh^2 b \sinh^2 a}{2 \cosh^2 b} \\ &= \frac{\frac{9 \cosh d - 7}{2(\cosh d - 1)9(\cosh d - 1)} - \frac{2(2 \cosh d - 1)}{2(\cosh d - 1)9(\cosh d - 1)}}{\frac{2 \cosh d - 1}{\cosh d - 1}} \\ &= \frac{5}{18} \frac{1}{2 \cosh d - 1} = \frac{5}{18} \frac{1}{4 \sinh^2 \frac{1}{2}d + 1} = \frac{5}{18} \frac{\coth^2 \frac{1}{2}d - 1}{\coth^2 \frac{1}{2}d + 3} \\ &> \frac{1}{4} \frac{\coth^2 \frac{1}{2}d - 1}{\coth^2 \frac{1}{2}d + 3} = \frac{1}{4} \frac{C(f, g) - 1}{C(f, g) + 3}. \end{aligned}$$

Taking the inverse hyperbolic sine yields

$$2(b - a) > 2 \log \left(\frac{1}{4} \frac{C(f, g) - 1}{C(f, g) + 3} + \sqrt{\left(\frac{1}{4} \frac{C(f, g) - 1}{C(f, g) + 3} \right)^2 + 1} \right) > 2 \log \left(\frac{1}{4} \frac{C(f, g) - 1}{C(f, g) + 3} + 1 \right),$$

and finally, since $\log(x + 1) > x/(x + 1)$ for $x > 0$, we have

$$2(b - a) > 2 \frac{\frac{1}{4} \frac{C(f, g) - 1}{C(f, g) + 3}}{\frac{1}{4} \frac{C(f, g) - 1}{C(f, g) + 3} + 1} = 2 \frac{C(f, g) - 1}{5 C(f, g) + 11} > \frac{1}{5} \frac{C(f, g) - 1}{C(f, g) + 3}.$$

For the second part of the theorem, suppose that

$$\tau(f), \tau(g) > \log C(f, g) + \frac{3}{2}.$$

Let $b' \geq 0$ be such that

$$\sinh b' = \sqrt{\frac{2}{\cosh d - 1}} = \frac{1}{\sinh \frac{1}{2}d},$$

and note that $h(b', b') = 1$ (see Figure 4.6). Lemma 4.17 implies that if $\tau(f), \tau(g) > 2b'$, then $\langle f, g \rangle$ is a Schottky semigroup of rank two. The proof is complete upon observing that

$$2b' = 2 \log \left(\frac{1 + \cosh \frac{1}{2}d}{\sinh \frac{1}{2}d} \right) < 2 \log(2 \coth \frac{1}{2}d) = \log(4C(f, g)) < \log C(f, g) + \frac{3}{2}. \quad \square$$

Note that by using Lemma 4.17 in the second part of Theorem 4.18 we also obtained the following.

COROLLARY 4.19. *Let f, g be two hyperbolic transformations with $C(f, g) > 1$. Suppose that*

$$\tau(f), \tau(g) > \log C(f, g) + \frac{3}{2}.$$

Then there exist open intervals A_f, B_f, A_g, B_g in $\overline{\mathbb{R}}$, with pairwise disjoint closures, that satisfy the following properties: the intervals A_f, B_f are symmetric with respect to f and $f(B_f^c)$ is contained in A_f ; the intervals A_g, B_g are symmetric with respect to g and $g(B_g^c)$ is contained in A_g .

Let us now move on to the case of two hyperbolic transformations with crossing axes. We start with a general lemma about limit sets of two-generator semigroups.

LEMMA 4.20. *Suppose that f and g are transformations in $\text{PSL}(2, \mathbb{R})$, with $f(x) = x$ and $g(y) = y$, for some points $x < y$ in $\overline{\mathbb{R}}$. Also, assume that the open interval (x, y) is mapped strictly inside itself by f and g . Then the forward limit set of $\langle f, g \rangle$ is the closed interval $[x, y]$ if and only if $g(x) \leq f(y)$.*

PROOF. Note that since f and g map (x, y) strictly inside itself, they are either parabolic or hyperbolic with attracting fixed points in $\{x, y\}$. We define $I = [x, y]$, and note that I is invariant under $\langle f, g \rangle$, which implies that the forward limit set of $\langle f, g \rangle$ is contained in I . If $f(y) < g(x)$, then the intervals $f(I)$ and $g(I)$ are disjoint and $\Lambda^+(\langle f, g \rangle) \subset f(I) \cup g(I)$, which implies that $\Lambda^+(\langle f, g \rangle)$ is a proper subset of I .

For the converse, assume that $g(x) \leq f(y)$. Then $g(I) \cup f(I) = I$, and thus for every $c \in I$ there exists $f_1 \in \{f, g\}$ such that $c \in f_1(I)$. So we can recursively find a sequence (f_n) with $f_n \in \{f, g\}$, and such that $c \in f_1 \circ f_2 \circ \dots \circ f_n(I)$, for all $n = 1, 2, \dots$. It is easy to check that the intervals $f_1 \circ f_2 \circ \dots \circ f_n(I)$ are nested and their Euclidean length converges

to 0 as $n \rightarrow \infty$. Hence, every $c \in I$ is an accumulation point of either x or y under $\langle f, g \rangle$, which implies that $\Lambda^+(\langle f, g \rangle) = I$. \square

Using Lemma 4.20 we can now prove the following analogue of Theorem 4.6 for two hyperbolic transformations with crossing axes.

THEOREM 4.21. *Suppose that f and g are hyperbolic transformations whose axes cross at an angle $\theta \in (0, \pi)$ and suppose that $\beta(g) < \alpha(f) < \alpha(g) < \beta(f)$.*

(1) *If $\tau(f), \tau(g) < \frac{1}{5}$, then $\Lambda^+(\langle f, g \rangle) = [\alpha(f), \alpha(g)]$.*

(2) *If $\tau(f), \tau(g) > |\log|C(f, g)|| + \frac{3}{2}$, then there exist open intervals A_f, B_f, A_g, B_g in $\overline{\mathbb{R}}$, with pairwise disjoint closures, that satisfy the following properties: the intervals A_f, B_f are symmetric with respect to f , and $f(B_f^c)$ is contained in A_f ; the intervals A_g, B_g are symmetric with respect to g , and $g(B_g^c)$ is contained in A_g .*

PROOF. For convenience, we conjugate f and g by a version of the Cayley transform so that they act on the unit disc \mathbb{D} , their axes cross at the origin and the Euclidean diameter landing at i and $-i$ bisects θ . For points $z, w \in \partial\mathbb{D}$, define $[z, w]$ to be the closed arc of the unit circle running counter-clockwise from z to w .

Let ϕ_g be the angle between the Euclidean radius landing at $g(\alpha(f))$ and the axis of g (see Figure 4.7 on the left). Applying the hyperbolic sine and cosine laws [6, Theorem 7.10.1] on the triangle with vertices $0, g(0)$ and $g(\alpha(f))$, we obtain

$$\cosh \tau(g) = \frac{\cos \phi_g \cos(\pi - \theta) + 1}{\sin \phi_g \sin(\pi - \theta)} \quad \text{and} \quad \sinh \tau(g) = \frac{\cos \phi_g + \cos(\pi - \theta)}{\sin \phi_g \sin(\pi - \theta)},$$

and therefore

$$(4.3.4) \quad \cos \phi_g = \frac{\sinh \tau(g) + \cosh \tau(g) \cos \theta}{\cosh \tau(g) + \sinh \tau(g) \cos \theta}.$$

Defining ϕ_f similarly and carrying out the same computations we can see that equation (4.3.4) holds if g is replaced by f .

Observe that if $\phi_f, \phi_g > \frac{\theta}{2}$ then $g(\alpha(f)) < f(\alpha(g))$ and so Lemma 4.20 implies that $\Lambda^+(\langle f, g \rangle) = [\alpha(f), \alpha(g)]$. So, it suffices to prove that if $\tau(f), \tau(g) < \frac{1}{5}$, then $\phi_f, \phi_g > \frac{\theta}{2}$. Note that these last inequalities for ϕ_f and ϕ_g are equivalent to $\cos \phi_f, \cos \phi_g < \cos \frac{\theta}{2}$, which

by substituting (4.3.4) (and the respective equation for $\cos \phi_f$) yield $\sinh \tau(f), \sinh \tau(g) < \varepsilon$, where

$$\varepsilon = \frac{\cos \frac{\theta}{2} - \cos \theta}{\sin \frac{\theta}{2} \sin \theta}.$$

Therefore, in order to complete the proof of the first part, it suffices to show that $\frac{1}{5} < \operatorname{arsinh} \varepsilon$, which follows from the string of inequalities below:

$$\begin{aligned} \operatorname{arsinh} \varepsilon &= \log \left(\frac{\cos \frac{\theta}{2} - \cos \theta}{\sin \frac{\theta}{2} \sin \theta} + \sqrt{\left(\frac{\cos \frac{\theta}{2} - \cos \theta}{\sin \frac{\theta}{2} \sin \theta} \right)^2 + 1} \right) \\ &= \log \left(\frac{\cos \frac{\theta}{2} - \cos \theta}{\sin \frac{\theta}{2} \sin \theta} + \frac{\sqrt{(1 - \cos \frac{\theta}{2} \cos \theta)^2}}{\sin \frac{\theta}{2} \sin \theta} \right) = \log \frac{(1 + \cos \frac{\theta}{2})(1 - \cos \theta)}{\sin \frac{\theta}{2} \sin \theta} \\ &= \log \left(\sqrt{2} \frac{(1 + \cos \frac{\theta}{2})(1 - \cos \theta)}{\sqrt{1 - \cos \theta} \sin \theta} \right) = \log \left(\sqrt{2} \frac{(1 + \cos \frac{\theta}{2})\sqrt{1 - \cos \theta}}{\sin \theta} \right) \\ &= \log \left(\sqrt{2} \frac{\left(1 + \sqrt{\frac{1 + \cos \theta}{2}}\right) \sqrt{1 - \cos \theta}}{\sin \theta} \right) = \log \frac{\sqrt{2} \sqrt{1 - \cos \theta} + \sqrt{1 - \cos^2 \theta}}{\sin \theta} \\ &= \log \frac{2 \sin \frac{\theta}{2} + \sin \theta}{\sin \theta} = \log \left(1 + \frac{1}{\cos \frac{\theta}{2}} \right) > \log 2 > \frac{1}{5}. \end{aligned}$$

For the second part of the theorem, recall that the diameters landing at $i, -i$ and $1, -1$, respectively, bisect the two complementary angles between the axes of f and g . Let σ_g be the angle between the axis of g and the radius landing at $g(-1)$ (see Figure 4.7 on the right). We are going to prove that if $\tau(g), \tau(f) > |\log|C(f, g)|| + \frac{3}{2}$, then g maps the complement of $[i, -1]$ in $\partial \mathbb{D}$ strictly inside $[-i, 1]$, and f maps the complement of $[1, i]$ strictly inside $[-1, -i]$. The proof will be carried out for g ; the situation for f is identical. It suffices to prove that $g(-1), g(i) \in (-i, 1)$. Considering the triangle with vertices $0, g(-1)$ and $g(0)$ and carrying out the same calculations we did for the first part, we obtain

$$\cos \sigma_g = \frac{\sinh \tau(g) - \cosh \tau(g) \cos \left(\frac{\pi}{2} - \frac{\theta}{2} \right)}{\cosh \tau(g) - \sinh \tau(g) \cos \left(\frac{\pi}{2} - \frac{\theta}{2} \right)}.$$

It is obvious that if $\sigma_g < \frac{\theta}{2}$ then $g(-1) \in (-i, 1)$. So substituting the equation above to $\cos \sigma_g > \cos \frac{\theta}{2}$ and solving for $\sinh \tau_g$ results in the inequality $\sinh \tau(g) > M$, where

$$M = \frac{\sin \frac{\theta}{2} + \cos \frac{\theta}{2}}{\sin \frac{\theta}{2} \cos \frac{\theta}{2}}.$$

One can easily check that working similarly in the triangle with vertices $0, g(0)$ and $g(i)$ yields the same estimate for the translation length of g .

So, in order to complete the proof, it suffices to prove that $|\log|C(f, g)|| + \frac{5}{2} > \operatorname{arsinh} M$. Observe that $M < 4/\sin \theta$, and from Lemma 4.15 we have $\sin \theta = 2\sqrt{-C(f, g)}/(1 - C(f, g))$. Thus

$$\begin{aligned} \operatorname{arsinh} M &< \log \left(\frac{4}{\sin \theta} + \sqrt{\frac{16}{\sin^2 \theta} + 1} \right) < \log \left(\frac{4 + \sqrt{17}}{\sin \theta} \right) < \log \left(5 \frac{1 - C(f, g)}{\sqrt{-C(f, g)}} \right) \\ &\leq \frac{1}{2} |\log -C(f, g)| + \frac{\log 10}{2} < |\log|C(f, g)|| + \frac{3}{2}, \end{aligned}$$

as required. \square

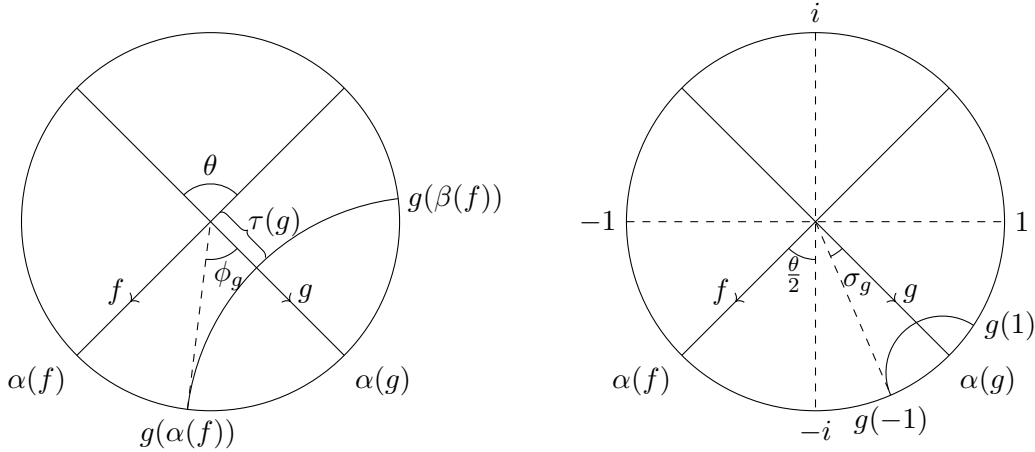


FIGURE 4.7. Two generators with crossing axes.

COROLLARY 4.22. *Let f and g be hyperbolic transformations, whose axes cross at an angle $\theta \in (0, \pi)$ and suppose that $\beta(g) < \alpha(f) < \alpha(g) < \beta(g)$. Also, assume that h is a hyperbolic transformation whose repelling fixed point lies in $(\alpha(f), \alpha(g))$. If $\tau(f), \tau(g) < \frac{1}{5}$, then the semigroup generated by f, g and h is not semidiscrete.*

PROOF. Assume that $\tau(f), \tau(g) < \frac{1}{5}$. Then from Theorem 4.21 we have that $\Lambda^+(\langle f, g \rangle) = [\alpha(f), \alpha(g)]$, and we can apply Lemma 4.12 to the semigroup $\langle f, g, h \rangle$ in order to deduce that it is either a discrete group or else it is not semidiscrete. Our goal is to prove that if $\tau(f), \tau(g) < \frac{1}{5}$, then $\langle f, g, h \rangle$ is not a discrete group.

Recall [6, Theorem 11.6.8], which states that if the group generated by f and g is discrete, then

$$\sinh \frac{1}{2}\tau(f) \sinh \frac{1}{2}\tau(g) \sin \theta \geq \cos \frac{3\pi}{7} \approx 0.223.$$

Observe that if

$$(4.3.5) \quad \sinh \frac{1}{2}\tau(f), \sinh \frac{1}{2}\tau(g) < \sqrt{\frac{\cos \frac{3\pi}{7}}{\sin \theta}},$$

then $\sinh \frac{1}{2}\tau(f) \sinh \frac{1}{2}\tau(g) \sin \theta < \cos \frac{3\pi}{7}$ and thus f, g do not generate a discrete group.

Inequality (4.3.5) is equivalent to $\sinh \tau(f), \sinh \tau(g) < \lambda$, where

$$\lambda = 2 \frac{\sqrt{\cos \frac{3\pi}{7} (\cos \frac{3\pi}{7} + \sin \theta)}}{\sin \theta}.$$

Hence, it suffices to prove that $\frac{1}{5} < \operatorname{arsinh} \lambda$. To that end, observe that $\lambda > 2 \cos \frac{3\pi}{7} > \frac{2}{5}$ and thus

$$\operatorname{arsinh} \lambda > \operatorname{arsinh} \frac{2}{5} = \log \left(\frac{2 + \sqrt{29}}{5} \right) > \frac{1}{5}. \quad \square$$

4.4. Proof of Theorem 4.6

We are now ready to prove Theorem 4.6. Suppose that f_1, f_2, \dots, f_n are hyperbolic Möbius transformations with distinct fixed points and $S = \langle f_1, f_2, \dots, f_n \rangle$ is not a Schottky semigroup of rank one. These assumptions imply that we cannot partition $\overline{\mathbb{R}}$ into two intervals I_1, I_2 , such that $\alpha_i \in I_1$ and $\beta_i \in I_2$, for all $i = 1, 2, \dots, n$. Hence, we can always find a pair of generators f_k and f_m such that either the axes of f_k and f_m are disjoint and $C(f_k, f_m) > 1$, or else their axes cross and there exists another generator f_l with $\alpha(f_k) < \beta(f_l) < \alpha(f_m)$, after a suitable conjugation. So, the first part of Theorem 4.6 comes as an application of the first part of Theorem 4.18 and Corollary 4.22.

Let us now focus on part (ii). For the rest of the proof we define $C_{i,j} = C(f_i, f_j)$. We are going to prove that if the translation lengths $\tau(f_k)$ are big enough, for all $k = 1, 2, \dots, n$, then there exists a finite collection of open intervals in $\overline{\mathbb{R}}$, with disjoint closures, whose

union is mapped compactly inside itself under each f_k .

Note that since the fixed points of the generators are distinct, $C_{i,j} \neq 0$ for all pairs i, j . Fix some $k \in \{1, 2, \dots, n\}$. Applying Corollary 4.19 and the second part of Theorem 4.21 to all pairs f_i, f_k with either $C_{i,k} < 0$ or $C_{i,k} > 1$, yields constants $M_{i,k} = |\log|C_{i,k}|| + \frac{3}{2}$ with the following properties. For each pair i, k , the inequalities $\tau(f_k) > M_{i,k}$ imply that there exist open intervals $A_{i,k}$ and $B_{i,k}$, symmetric with respect to f_k , such that $\alpha_k \in A_{i,k}$, $\beta_k \in B_{i,k}$ and f_k maps the complement of $B_{i,k}$ inside $A_{i,k}$. In addition, due to the way the intervals $A_{i,k}, B_{i,k}$ were constructed in Corollary 4.19 (Lemma 4.17 to be more precise) and the second part of Theorem 4.21, we can choose $A_{i,k}$ and $B_{i,k}$ to have the following properties: if δ_A is the hyperbolic geodesic with the same endpoints as $A_{i,k}$ and δ_B is the hyperbolic geodesic with the same endpoints as $B_{i,k}$, then the hyperbolic distance $\rho(\delta_A, \delta_B)$ between δ_A and δ_B is $M_{i,k}$. From the symmetry of the intervals with respect to f_k , the collections $\{A_{i,k}\}_i$ and $\{B_{i,k}\}_i$ consist of nested intervals. Let $A^k \in \{A_{i,k}\}_i$ and $B^k \in \{B_{i,k}\}_i$ be the innermost interval in each collection. Note that $M_{i,j}, A_{i,j}, B_{i,j}$ are symmetric with respect to the indices i, j .

Because $A_{i,k} \cap B_{i,k} = \emptyset$ for $i = 1, 2, \dots, n$, the intervals A^k and B^k are disjoint. Note that it is possible that $A^k = A_{j,k}$ and $B^k = B_{l,k}$ for different pairs j, k and k, l . Let ℓ_A be the hyperbolic line in \mathbb{H} that has the same endpoints as A^k and ℓ_B the hyperbolic line with the same endpoints as B^k . Also, denote the distance $\rho(\ell_A, \ell_B)$ between these two lines by d_k . Recall that the lines ℓ_A and ℓ_B have to be perpendicular to the axis of f_k . It is easy to see that if $\tau(f_k) > d_k$ then f_k maps the complement of $\overline{B^k}$ inside A^k . We claim that $M > d_k$, where

$$(4.4.1) \quad M = 2 \max_{m,n} \left\{ |\log|C_{m,n}|| + \frac{3}{2} \right\} + \max_{m,n} \{ \rho(\text{Ax}(f_m), \text{Ax}(f_n)) \},$$

and both maxima are taken over all pairs f_m, f_n (recall that $C_{m,n} \neq 0$ for all m, n).

If $d_k < M_{j,k} + M_{l,k}$, our claim is obvious because $M_{j,k} + M_{l,k} < 2 \max\{M_{m,n}\}$. Assume that $d_k \geq M_{j,k} + M_{l,k}$, and let γ_B be the hyperbolic line with the same endpoints as $B_{j,k}$ and γ_A the line with the same endpoints as $A_{l,k}$ (see Figure 4.8). Recall that γ_B and γ_A are perpendicular to the axes of f_k . Also, as we noted in the beginning of the proof, we have that $\rho(\ell_A, \gamma_B) = M_{j,k}$ and $\rho(\ell_B, \gamma_A) = M_{l,k}$. In other words, γ_B is the unique hyperbolic line, perpendicular to $\text{Ax}(f_k)$ that is a distance $M_{j,k}$ away from ℓ_A and closer

to ℓ_B . So, $d_k = M_{j,k} + M_{l,k} + \rho(\gamma_B, \gamma_A)$. Consider the hyperbolic half-plane $H(\gamma_B)$ that is bounded by γ_B and contains ℓ_A , and define $H(\gamma_A)$ similarly. Because $d_k \geq M_{j,k} + M_{l,k}$, these two half-planes are disjoint. Also, since A^k and B^k were chosen to be the innermost intervals in their respective collection, for the transformations f_j and f_l we have that $\text{Ax}(f_j) \subset H(\gamma_B)$ and $\text{Ax}(f_l) \subset H(\gamma_A)$. Hence, the axes of f_j and f_l are disjoint, which implies that $\rho(\gamma_B, \gamma_A) < \rho(\text{Ax}(f_j), \text{Ax}(f_l))$, proving our claim.

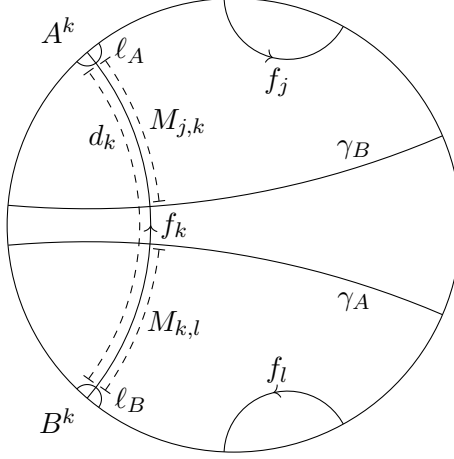


FIGURE 4.8. The intervals A^k and B^k for a generator f_k .

Since k was chosen arbitrarily, if $\tau(f_k) > M$ for all $k = 1, 2, \dots, n$, then for each k there exist open intervals A^k, B^k , with disjoint closures, such that f_k maps the complement of B^k inside A^k . Also, as all the fixed points of the generators are distinct, the collection $\{A^k, B^k : k = 1, 2, \dots, n\}$ consists of disjoint intervals, and thus each f_k maps $\bigcup_i A^i$ compactly inside A^k , which implies that $\langle f_1, f_2, \dots, f_n \rangle$ is a Schottky semigroup. In order to complete the proof, it suffices to show that

$$M \leq 4 \max_{m,n} \left\{ |\log|C_{m,n}|| + \log \frac{\sqrt{|C_{m,n}|} + 1}{|\sqrt{|C_{m,n}|} - 1|} \right\} + 3,$$

where the maximum is taken over all pairs f_m, f_n , with $C(f_m, f_n) \neq -1$.

To that end, note that both maxima in M can be taken over all pairs f_m, f_n , with $C(f_m, f_n) \neq -1$, because if $C(f_i, f_j) = -1$, for some pair f_i, f_j , then

$$|\log|C_{i,j}|| + \frac{3}{2} + \rho(\text{Ax}(f_i), \text{Ax}(f_j)) = \frac{3}{2} \leq M.$$

Now, recall that if $\text{Ax}(f_m)$ and $\text{Ax}(f_n)$ are disjoint (i.e. $C_{m,n} > 0$) then

$$\cosh \rho(\text{Ax}(f_m), \text{Ax}(f_n)) = \frac{C_{m,n} + 1}{|C_{m,n} - 1|},$$

and so

$$\rho(\text{Ax}(f_m), \text{Ax}(f_n)) = \log \frac{\sqrt{C_{m,n}} + 1}{|\sqrt{C_{m,n}} - 1|}.$$

Therefore,

$$M \leq 4 \max_{m,n} \left\{ |\log |C_{m,n}| | + \log \frac{\sqrt{|C_{m,n}|} + 1}{|\sqrt{|C_{m,n}|} - 1|} \right\} + 3,$$

and the maximum is taken over all pairs f_m, f_n with $C_{m,n} \neq -1$. □

CHAPTER 5

Möbius semigroups and uniform hyperbolicity

5.1. The hyperbolic locus

The rest of this thesis will be dedicated to exploring the interactions between semigroups of Möbius transformations and $\mathrm{PSL}(2, \mathbb{R})$ -cocycles.

For a fixed positive integer N , we consider the parameter space $\mathrm{PSL}(2, \mathbb{R})^N$. The set $\mathrm{PSL}(2, \mathbb{R})^N$ can be thought of as a subspace of the Euclidean space \mathbb{R}^{3N} and so it carries a natural topology. Recall that $\mathrm{PSL}(2, \mathbb{R})$ also carries the topology of locally uniform convergence. On several occasions throughout these last two chapters it will be convenient to consider $\mathrm{PSL}(2, \mathbb{R})^N$ imbued with the product topology on the group $\mathrm{PSL}(2, \mathbb{R})$ with the topology of locally uniform convergence.

DEFINITION 5.1. We define \mathcal{H} to be the set of N -tuples (f_1, f_2, \dots, f_N) in $\mathrm{PSL}(2, \mathbb{R})^N$ for which there exists a finite union M of open intervals in $\overline{\mathbb{R}}$, with disjoint closures, such that each f_i maps \overline{M} into M . The parameter space \mathcal{H} is called the *hyperbolic locus*.

From now on, and for the rest of this thesis, we shall use \mathcal{F} to denote an N -tuple in $\mathrm{PSL}(2, \mathbb{R})^N$. An N -tuple \mathcal{F} that lies in \mathcal{H} will be called *uniformly hyperbolic*, and the set M in Definition 5.1 is called a *multicone* of \mathcal{F} .

The hyperbolic locus was studied by Avila, Bochi and Yoccoz [2], who proved that \mathcal{H} is exactly the set of N -tuples that correspond to a uniformly hyperbolic $\mathrm{PSL}(2, \mathbb{R})$ -cocycle over the full shift on N symbols (see [2, Theorem 2.2]), which explains our notation in the previous paragraph. The concept of uniform hyperbolicity plays an important role in the theory of cocycle dynamics due to its connection with various other aspects of dynamical systems, such as Lyapunov exponents [13] and dimensions of self-affine sets [5].

The goal of this chapter is to investigate the structure of the hyperbolic locus by studying semigroups arising from the N -tuples in \mathcal{H} , or $\mathrm{PSL}(2, \mathbb{R})^N$ in general. To that end, we associate a semigroup of Möbius transformations to each N -tuple \mathcal{F} in the following way.

DEFINITION 5.2. For $\mathcal{F} = (f_1, f_2, \dots, f_N)$ in $\mathrm{PSL}(2, \mathbb{R})^N$, we define the *semigroup generated by \mathcal{F}* , which we will denote by $\langle \mathcal{F} \rangle$, to be the semigroup generated by $\{f_1, f_2, \dots, f_N\}$.

It is easy to check that if \mathcal{F} lies in \mathcal{H} then $\langle \mathcal{F} \rangle$ is a Schottky semigroup (see Definition 4.5). In addition, since the semigroup $\langle \mathcal{F} \rangle$ maps an open subset of $\overline{\mathbb{R}}$ compactly inside itself, it is *purely hyperbolic*, i.e., it only contains hyperbolic transformations. Note that for any N -tuple \mathcal{F} , the semigroup $\langle \mathcal{F} \rangle$ is finitely generated. In the following lemma we prove an elementary, yet useful, property of finitely generated semigroups.

LEMMA 5.3. *Let $\mathcal{F} = (f_1, f_2, \dots, f_N)$ be an N -tuple in $\mathrm{PSL}(2, \mathbb{R})^N$ and suppose that (h_n) is a sequence in $\langle \mathcal{F} \rangle$, which does not have any constant subsequences. Then there exists a subsequence (h_{n_l}) of (h_n) , such that $h_{n_l} = G_l k_l$, for all positive integer l , where (G_l) is a right-composition sequence in $\langle \mathcal{F} \rangle$ and k_l lies in $\langle \mathcal{F} \rangle$ for all l .*

PROOF. Since each element of (h_n) is a composition of elements from $\{f_1, f_2, \dots, f_N\}$, we can find a subsequence (h_{n_m}) of (h_n) such that $h_{n_m} = g_1 \phi_m$, for all positive integers m , where $g_1 \in \{f_1, f_2, \dots, f_N\}$ and (ϕ_m) is a sequence in $\langle \mathcal{F} \rangle$. Repeating this argument for the sequence (ϕ_m) , we obtain a subsequence (ϕ_{m_ν}) of (ϕ_m) , such that $\phi_{m_\nu} = g_2 \psi_\nu$, for all ν , where $g_2 \in \{f_1, f_2, \dots, f_N\}$ and (ψ_ν) is a sequence in S . Hence, $h_{n_{m_\nu}} = g_1 g_2 \psi_\nu$, for all positive integers ν . Iterating this process yields the desired result. \square

In order to understand the properties of semigroups generated by uniformly hyperbolic N -tuples, we will now describe several results from [2] and [46] using the theory of Möbius semigroups that we have established so far.

For an N -tuple \mathcal{F} in $\mathrm{PSL}(2, \mathbb{R})^N$, we will denote by $\Lambda^+(\mathcal{F})$ and $\Lambda^-(\mathcal{F})$ the forward and backward limit sets of the semigroup $\langle \mathcal{F} \rangle$, respectively.

DEFINITION 5.4. Let $\mathcal{F} \in \mathrm{PSL}(2, \mathbb{R})^N$. The complement of the union of the connected components of $\overline{\mathbb{R}} \setminus \Lambda^+(\mathcal{F})$ that intersect $\Lambda^-(\mathcal{F})$ will be called *the forward core* of \mathcal{F} and will be denoted by $C^+(\mathcal{F})$. Similarly, we define the *backward core* $C^-(\mathcal{F})$ of \mathcal{F} to be the complement of the union of the connected components of $\overline{\mathbb{R}} \setminus \Lambda^-(\mathcal{F})$ that intersect $\Lambda^+(\mathcal{F})$.

Note that, for any $\mathcal{F} \in \mathrm{PSL}(2, \mathbb{R})^N$, because $\Lambda^+(\mathcal{F})$ and $\Lambda^-(\mathcal{F})$ are closed sets, the cores $C^+(\mathcal{F})$ and $C^-(\mathcal{F})$ are also closed. Also, for a generic N -tuple \mathcal{F} , the cores $C^+(\mathcal{F})$

and $C^-(\mathcal{F})$ may be trivial subsets of the extended real line. If \mathcal{F} generates a Schottky semigroup, then we have the next lemma. It is of interest to observe that for any $\mathcal{F} \in \mathrm{PSL}(2, \mathbb{R})^N$, the components of $C^+(\mathcal{F})$ and $C^-(\mathcal{F})$ appear in alternating order around $\overline{\mathbb{R}}$. Therefore, $C^+(\mathcal{F})$ is finitely connected if and only if $C^-(\mathcal{F})$ is finitely connected. See Figure 5.1 for examples of the forward and backward cores of Schottky semigroups.

LEMMA 5.5. *Let $\langle \mathcal{F} \rangle$ be a Schottky semigroup. The forward and backward cores of \mathcal{F} satisfy the following properties:*

- (1) $C^+(\mathcal{F}) \cap C^-(\mathcal{F}) = \partial C^+(\mathcal{F}) \cap \partial C^-(\mathcal{F})$;
- (2) $\partial C^+(\mathcal{F}) \subset \Lambda^+(\mathcal{F})$ and $\partial C^-(\mathcal{F}) \subset \Lambda^-(\mathcal{F})$;
- (3) $\Lambda^+(\mathcal{F}) \subseteq C^+(\mathcal{F})$ and $\Lambda^-(\mathcal{F}) \subseteq C^-(\mathcal{F})$;
- (4) $C^+(\mathcal{F})$ and $C^-(\mathcal{F})$ are finitely connected;
- (5) $f(C^+(\mathcal{F})) \subseteq C^+(\mathcal{F})$ and $f^{-1}(C^-(\mathcal{F})) \subseteq C^-(\mathcal{F})$, for all $f \in \langle \mathcal{F} \rangle$.

PROOF. Since $\langle \mathcal{F} \rangle$ is a Schottky semigroup, there exists a finite union U of open intervals, with disjoint closures, that is mapped strictly inside itself under $\langle \mathcal{F} \rangle$ (see Definition 4.5). So, $\Lambda^+(\mathcal{F})$ lies in \overline{U} and $\Lambda^-(\mathcal{F})$ lies in the complement of U . Using similar arguments, we can show that there exists a finite union of open intervals V , disjoint from U , such that $\langle \mathcal{F} \rangle^{-1}$ maps V strictly inside itself. Thus $C^+(\mathcal{F}) \subseteq \overline{U}$, and $C^-(\mathcal{F}) \subseteq \overline{V} \subseteq U^c$. Parts (1) to (4) follow easily from these facts.

We are now going to show that $C^+(\mathcal{F})$ is mapped inside itself by $\langle \mathcal{F} \rangle$. The proof for $C^-(\mathcal{F})$ is similar. Let I be a connected component of $C^+(\mathcal{F})$. Suppose that there exists an ordinate f of \mathcal{F} such that $f(I)$ is not contained in $C^+(\mathcal{F})$, and note that parts (2) and (3) imply that the endpoints of I are mapped into $C^+(\mathcal{F})$. Hence, there exist points $x \in I$ and $y \in \Lambda^-(\mathcal{F})$ such that $f(x) = y$. Therefore $x = f^{-1}(y) \in \Lambda^-(\mathcal{F})$, which contradicts parts (1) and (3) of this lemma. So all connected components of $C^+(\mathcal{F})$ are mapped into $C^+(\mathcal{F})$ by each ordinate of \mathcal{F} , and so we obtain the desired result. \square

We now provide a characterisation of uniformly hyperbolic N -tuples in terms of the limits sets of the semigroups they generate. This result can be easily inferred from the material in [2]; we provide the proof for the sake of completeness.

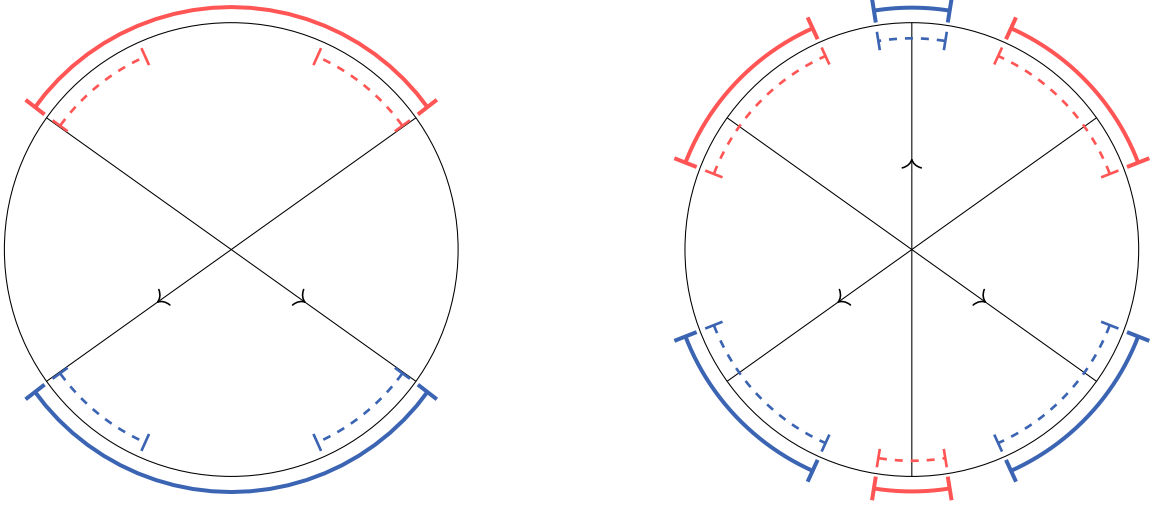


FIGURE 5.1. Examples of cores. The dashed lines indicate the limit sets, while the continuous lines the cores.

LEMMA 5.6 ([2]). *An N -tuple \mathcal{F} is uniformly hyperbolic if and only if all ordinates of \mathcal{F} are hyperbolic transformations and $\Lambda^+(\mathcal{F}) \cap \Lambda^-(\mathcal{F}) = \emptyset$.*

PROOF. Suppose that \mathcal{F} is uniformly hyperbolic, and let M be its multicone. Recall that uniform hyperbolicity implies that $\langle \mathcal{F} \rangle$ only contains hyperbolic transformations. Since M is mapped compactly inside itself, all the attracting fixed points of elements of $\langle \mathcal{F} \rangle$ lie in M . Also, M does not contain any repelling fixed points of elements of $\langle \mathcal{F} \rangle$. Hence, as $\Lambda^+(\mathcal{F})$ and $\Lambda^-(\mathcal{F})$ are the closures of the attracting and repelling fixed points of elements of $\langle \mathcal{F} \rangle$, respectively, $\Lambda^+(\mathcal{F})$ is contained in M and does not intersect $\Lambda^-(\mathcal{F})$. For the converse, note that because all the ordinates of \mathcal{F} are hyperbolic and the limit sets are disjoint, $\langle \mathcal{F} \rangle$ has to be inverse-free. Using arguments similar to the proof of Lemma 5.5, it is easy to check that $\Lambda^+(\mathcal{F}) \cap \Lambda^-(\mathcal{F}) = \emptyset$ implies that $C^+(\mathcal{F}) \cap C^-(\mathcal{F}) = \emptyset$ and the sets $C^+(\mathcal{F})$ and $C^-(\mathcal{F})$ are forward and backward invariant, respectively. Thus [2, Lemma 2.7] is applicable and yields that \mathcal{F} is uniformly hyperbolic. \square

Suppose that $\{f_1, f_2, \dots, f_N\}$ is a finite collection of transformations from $\text{PSL}(2, \mathbb{R})$. Then for any positive integer m , we define $W_m(f_1, f_2, \dots, f_N)$ to be the collection of all possible compositions of m elements from $\{f_1, f_2, \dots, f_N\}$. That is,

$$W_m(f_1, f_2, \dots, f_N) = \{g_1 g_2 \cdots g_m : g_i \in \{f_1, f_2, \dots, f_N\}, \text{ for all } i = 1, 2, \dots, m\}.$$

The set $W_m(f_1, f_2, \dots, f_N)$ is essentially the collection of all words of length m that arise from the alphabet $\{f_1, f_2, \dots, f_N\}$, which justifies our notation.

LEMMA 5.7. *Suppose that $\{f_1, f_2, \dots, f_N\}$ is a finite collection in $\mathrm{PSL}(2, \mathbb{R})$ and m a positive integer. The forward limit set of the semigroup generated by $W_m(f_1, f_2, \dots, f_N)$ is $\Lambda^+(\langle f_1, f_2, \dots, f_N \rangle)$ and the backward limit set of the semigroup generated by $W_m(f_1, f_2, \dots, f_N)$ is $\Lambda^-(\langle f_1, f_2, \dots, f_N \rangle)$.*

PROOF. The proof will be carried out for the forward limit sets; the case for the backward limit sets is similar. We denote by S_m the semigroup generated by $W_m(f_1, f_2, \dots, f_N)$, and by S the semigroup $\langle f_1, f_2, \dots, f_N \rangle$. Observe that we only need to show that $\Lambda^+(S) \subseteq \Lambda^+(S_m)$, since the other inclusion follows from the fact that S_m is contained in S .

Suppose that x lies in $\Lambda^+(S)$. Then there exists a sequence (h_n) in S , such that $h_n(i)$ converges to x . Write each element of (h_n) as a word in the alphabet $\{f_1, f_2, \dots, f_N\}$, and denote by $\ell(n)$ the length of the word h_n , for all n . There exist integers q_n and $r_n \in \{0, 1, \dots, m-1\}$ such that $\ell(n) = q_n m + r_n$, for all $n \in \mathbb{N}$. Hence, we can write h_n as

$$h_n = g_{n_1} g_{n_2} \dots g_{n_m} g_{n_{m+1}} g_{n_{m+2}} \dots g_{n_{2m}} \dots g_{n_{q_n m}} \phi_{n_1} \phi_{n_2} \dots \phi_{n_{r_n}},$$

where g_{n_i} and ϕ_{n_j} lie in $\{f_1, f_2, \dots, f_N\}$, for all i, j . Since there are only finitely many choices for r_n and ϕ_{n_j} , we can assume, by passing to a subsequence of (h_n) , that

$$\phi_{n_1} \phi_{n_2} \dots \phi_{n_{r_n}} = \phi,$$

for some transformation $\phi \in S$ and all positive integers n . Therefore, for every positive integer n we can find a transformation k_n in S_m , such that $h_n(i) = k_n(\phi(i))$. So the sequence $(h_n(i))$ is contained in the orbit of the point $\phi(i)$ under the semigroup S_m , and since $(h_n(i))$ converges to x , we conclude that $x \in \Lambda^+(S_m)$. This implies that $\Lambda^+(S) \subseteq \Lambda^+(S_m)$, as required. \square

The final theorem of this section is a result of Yoccoz [46, Proposition 2], which states that uniform hyperbolicity is equivalent to the uniform linear growth of translation lengths for elements of $\langle \mathcal{F} \rangle$. In the following theorem, and for the rest of this thesis, we make the convention that elliptic and parabolic Möbius transformations have zero translation length.

THEOREM 5.8 ([46]). *An N -tuple $\mathcal{F} = (f_1, f_2, \dots, f_N)$ is uniformly hyperbolic if and only if there exists a constant $c > 0$ such that for all positive integers m we have*

$$\tau(g) \geq mc, \quad \text{for all } g \in W_m(f_1, f_2, \dots, f_N).$$

5.2. The semidiscrete and inverse-free locus

As we mentioned in the previous section, if an N -tuple \mathcal{F} is uniformly hyperbolic then all elements of $\langle \mathcal{F} \rangle$ are hyperbolic transformations. It is therefore natural to define the set \mathcal{E} of all N -tuples in $\mathrm{PSL}(2, \mathbb{R})^N$ for which $\langle \mathcal{F} \rangle$ contains elliptic elements. The parameter space \mathcal{E} plays an important role in the work of Yoccoz [46] and Avila, Bochi and Yoccoz [2] due to its interaction with the hyperbolic locus. In particular, these authors prove that the set \mathcal{E} is open and connected [2, Proposition A.3] and that $\mathcal{H}^c = \overline{\mathcal{E}}$ [46, Proposition 6], where the complement and closure are taken in $\mathrm{PSL}(2, \mathbb{R})^N$. In [2, Theorem 3.3] it is shown that when $N = 2$, the hyperbolic locus and \mathcal{E} share the same boundary, i.e. $\overline{\mathcal{H}} = \mathcal{E}^c$, and ask whether the same holds for any positive integer N (see [2, Question 4] and [46, Question 4]). This question was answered in the negative by Jacques and Short [26, Section 16], where for all $N > 2$ they constructed an N -tuple that lies neither in \mathcal{E} nor in the closure of \mathcal{H} . Their construction motivated a restatement of the original question of Avila, Bochi and Yoccoz which we now describe. First, we make the following definition.

DEFINITION 5.9. The parameter space \mathcal{E}_I of all N -tuples in $\mathrm{PSL}(2, \mathbb{R})^N$ for which $\langle \mathcal{F} \rangle$ contains elliptic elements or the identity is called the *elliptic locus*.

LEMMA 5.10. *If \mathcal{F} lies in the elliptic locus then the semigroup $\langle \mathcal{F} \rangle$ is not semidiscrete and inverse-free.*

PROOF. If $\langle \mathcal{F} \rangle$ contains the identity then it is not inverse-free. Assume that it contains an elliptic transformation h . If there exists a positive integer k such that $h^k = \mathrm{Id}$ then, again, $\langle \mathcal{F} \rangle$ is not inverse-free. Otherwise, we can find a sequence of integers (n_k) , so that (h^{n_k}) accumulates to the identity, and thus $\langle \mathcal{F} \rangle$ is not semidiscrete. \square

The question of Jacques and Short is the following:

QUESTION 5.11 ([26]). Is it true that $\overline{\mathcal{H}} = \mathcal{E}_I^c$ in $\mathrm{PSL}(2, \mathbb{R})^N$?

The rest of this chapter will be dedicated to investigating this question. Originally, Avila, Bochi and Yoccoz defined the elliptic locus to be set \mathcal{E} . However, for our purposes, using the set \mathcal{E}_I is far more convenient.

In order to tackle Question 5.11 we define a new parameter space in $\mathrm{PSL}(2, \mathbb{R})^N$.

DEFINITION 5.12. The set \mathcal{S} of all N -tuples \mathcal{F} in $\mathrm{PSL}(2, \mathbb{R})^N$ that generate a semidiscrete and inverse-free semigroup is called the *semidiscrete and inverse-free locus*.

Observe that the semidiscrete and inverse-free locus \mathcal{S} contains the hyperbolic locus since, as we mentioned, every uniformly hyperbolic N -tuple generates a Schottky semigroup. Recall that in Lemma 4.8 we described a semidiscrete and inverse-free semigroup whose limit sets are not disjoint, which indicates that the hyperbolic locus is strictly contained in \mathcal{S} (see Lemma 5.6). In fact, using Theorem 5.8 we can show that uniformly hyperbolic N -tuples generate discrete semigroups, which is not true in general for N -tuples in \mathcal{S} (see Example 4.2 and Lemma 4.8).

THEOREM 5.13. *If \mathcal{F} lies in \mathcal{H} , then $\langle \mathcal{F} \rangle$ is a discrete semigroup.*

PROOF. Suppose that \mathcal{F} lies in \mathcal{H} and there exists a sequence (f_n) in $\langle \mathcal{F} \rangle$ that converges to some $f \in \mathrm{PSL}(2, \mathbb{R})$. Then, the translation lengths $\tau(f_n)$ of f_n have to converge to $\tau(f)$, as $n \rightarrow \infty$, and Theorem 5.8 implies that f_n has to eventually be equal to f (recall that we write $\tau(f) = 0$ when f is parabolic). \square

It is easy to check that the converse of Theorem 5.13 does not hold. If \mathcal{F} generates a semigroup that is a discrete group, then \mathcal{F} does not lie in \mathcal{H} since $\langle \mathcal{F} \rangle$ contains the identity. In the next chapter we are going to construct a discrete and inverse-free semigroup that does not lie in \mathcal{H} (see Section 6.3, to follow).

5.3. The complement of \mathcal{S}

We now aim to answer the main question of this chapter, Question 5.11, in the negative. We are going to give an example of an N -tuple that lies in the complement of \mathcal{E}_I , but not in the closure of \mathcal{H} . Note that because $\mathcal{H} \subset \mathcal{S}$, it suffices for our N -tuple to not lie in the

closure of \mathcal{S} .

First, we need to establish certain topological properties of \mathcal{S} . Note that from Lemma 5.10 the semidiscrete and inverse-free locus is disjoint from the elliptic locus. One can easily check that the hyperbolic locus is open in $\mathrm{PSL}(2, \mathbb{R})^N$, but it is not connected. One of the main objectives of [2] is to study the structure of the connected components of \mathcal{H} , called *the hyperbolic components*. For $N = 2$, the authors give a full description of the hyperbolic components, but note that for $N \geq 3$ “new phenomena appear, which make such a complete description much more difficult and complicated”. There is one component of \mathcal{H} , however, that is easy to describe, which we now define.

DEFINITION 5.14. The component of \mathcal{H} that contains all uniformly hyperbolic N -tuples that have a multicone consisting of a single interval is called *the principal component* of \mathcal{H} .

If \mathcal{F} has a multicone that consists of one open interval I , then the attracting fixed points of all the ordinates of \mathcal{F} lie in I and all their repelling fixed points in the complement of \bar{I} . Suppose that \mathcal{F}' has a multicone that consists of an open interval J . Then, we can continuously deform \mathcal{F} , in the space $\mathrm{PSL}(2, \mathbb{R})^N$, so that the interval I is deformed into J and the fixed points of the n^{th} ordinate of \mathcal{F} are deformed into the fixed points of the n^{th} ordinate of \mathcal{F}' . Thus the set of all N -tuples that map a nontrivial interval of $\bar{\mathbb{R}}$ compactly inside itself is connected and the principal component is well-defined. We shall denote the principal component by H_P and call all other components of \mathcal{H} *non-principal*.

In order to examine the properties of the principal component, we make the following definition.

DEFINITION 5.15. Let J be an open, nontrivial interval in $\bar{\mathbb{R}}$. We define $\mathcal{M}(J)$ to be the semigroup of Möbius transformations that map J inside itself.

Note that if \mathcal{F} generates a Schottky semigroup of rank one, then $\langle \mathcal{F} \rangle$ is contained in $\mathcal{M}(J)$, for some interval J (see Definition 4.5). The converse, however, is not true since semigroups in $\mathcal{M}(J)$ are not necessarily semidiscrete and inverse-free. As an easy example, suppose that (f, g) generates a Schottky semigroup of rank one. Then the semigroup $\langle f, g, \mathrm{Id} \rangle$ lies in $\mathcal{M}(J)$, for some J , but is not a Schottky semigroup.

LEMMA 5.16. *Suppose that $\mathcal{F} \in \mathrm{PSL}(2, \mathbb{R})^N$ is such that $\langle \mathcal{F} \rangle$ lies in $\mathcal{M}(J)$, for some open, nontrivial interval $J \subset \overline{\mathbb{R}}$. Then \mathcal{F} lies in the closure of the principal component.*

PROOF. Assume, without loss of generality, that $J = (0, 1)$. Observe that the transformations in $\mathcal{M}(J)$ are either hyperbolic transformations whose attracting fixed points lie in \overline{J} and repelling fixed points in J^c , parabolic transformations that fix one endpoint of J and map J inside itself, or the identity. Now, let U be a neighbourhood of \mathcal{F} in $\mathrm{PSL}(2, \mathbb{R})^N$. Observe that a parabolic transformation that fixes 0 or 1 can be approximated by a sequence of hyperbolic transformations (f_n) , with $\alpha(f_n) \in (0, 1)$ and $\beta(f_n) \in [0, 1]^c$. Therefore, we can find \mathcal{F}' in U such that all ordinates of \mathcal{F}' are hyperbolic transformations with attracting fixed points in J , and repelling fixed points outside \overline{J} . Hence, $\langle \mathcal{F}' \rangle$ maps J compactly inside itself and thus lies in H_P . \square

Lemma 5.16 implies that N -tuples that generate Schottky semigroups of rank one lie in the closure of H_P . It is also easy to see that all points in H_P generate Schottky semigroups of rank one. However, not all points on the boundary of the principal component generate semidiscrete and inverse-free semigroups. In order to see this, consider two hyperbolic transformations f and g , with $\mathrm{Ax}(f) = \mathrm{Ax}(g)$ and $\alpha(f) = \beta(g)$, and let $\mathcal{F} = (f, g)$. It is easy to check that $\langle \mathcal{F} \rangle$ is not semidiscrete and inverse-free but lies on the boundary of H_P , as it can be approximated by two hyperbolic transformations with crossing axes. Therefore, the semidiscrete and inverse-free locus is not a closed set. However, we can prove the following.

THEOREM 5.17. *The set $\mathcal{S} \setminus \overline{H_P}$ is closed in $\mathrm{PSL}(2, \mathbb{R})^N$.*

Theorem 5.17 indicates that H_P is special in the sense that it is the only hyperbolic component whose boundary contains N -tuples that do not generate semidiscrete and inverse-free semigroups. It also yields the following useful corollary.

COROLLARY 5.18. *The closure of \mathcal{S} in $\mathrm{PSL}(2, \mathbb{R})^N$ is $\mathcal{S} \cup \overline{H_P}$.*

In order to prove Theorem 5.17 we are going to need certain preliminary results. First, we have a version of Jørgensen's inequality for semigroups, which can be found in [26, Theorem 12.11].

THEOREM 5.19 ([26]). *Suppose that f and g are Möbius transformations in $\mathrm{PSL}(2, \mathbb{R})$ such that $\langle f, g \rangle$ is non-elementary and semidiscrete. Then either $\langle f, g \rangle$ is a Schottky semigroup of rank one, or else*

$$||\mathrm{tr}(f)|^2 - 4| + ||\mathrm{tr}([f, g])| - 2| \geq 1,$$

where $[f, g] = fgf^{-1}g^{-1}$.

Suppose that f and g are either hyperbolic or parabolic transformations in $\mathrm{PSL}(2, \mathbb{R})$, with no common fixed points. We recall the following definition from [26, Section 12] (see also [26, Lemma 12.2]).

DEFINITION 5.20. The pair of transformations f and g is called *antiparallel* if the semigroup $\langle f, g \rangle$ is not a Schottky semigroup of rank one.

Lemma 5.16 implies that if f and g are antiparallel then (f, g) does not lie in the closure of the principal component in $\mathrm{PSL}(2, \mathbb{R})^2$. Note that if both f and g are hyperbolic, then they are antiparallel if and only if $C(f, g) > 1$ (see Lemma 4.15). However, we cannot have a similar cross ratio condition in the case when either f or g is parabolic, since then $C(f, g)$ is always one. See Figure 5.2 for configurations of antiparallel pairs.

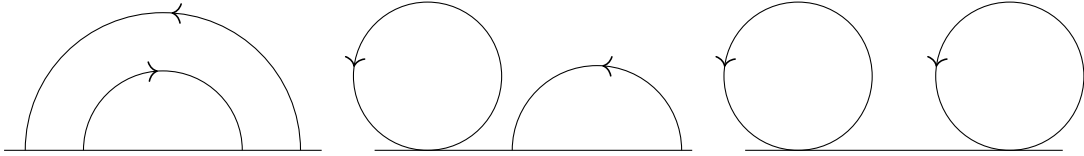


FIGURE 5.2. Pairs of antiparallel transformations.

We first prove an elementary lemma about the commutator of Möbius transformations.

LEMMA 5.21. *Let f and g be hyperbolic or parabolic transformations in $\mathrm{PSL}(2, \mathbb{R})$ with a common fixed point. Then, for any transformations h and k in $\langle f, g \rangle$ we have that $|\mathrm{tr}([h, k])| = 2$.*

PROOF. Conjugate f and g by a transformation in $\mathrm{PSL}(2, \mathbb{R})$ so that they both fix the point at infinity. Then, any transformation ϕ in $\langle f, g \rangle$ is of the form $\phi(z) = \lambda z + \kappa$, for some $\lambda > 0$ and $\kappa \in \mathbb{R}$. Suppose that $h(z) = az + b$ and $k(z) = cz + d$ lie in $\langle f, g \rangle$. Then we

can easily see that the commutator of h and k is given by $[h, k](z) = z + (a-1)d - (c-1)d$. Therefore, $[h, k]$ is either parabolic or the identity, which yields the desired result. \square

LEMMA 5.22. *Suppose that (f_n) and (g_n) are sequences of hyperbolic or parabolic transformations in $\mathrm{PSL}(2, \mathbb{R})$ such that f_n and g_n do not have any common fixed points and are antiparallel, for all positive integers n . Assume that (f_n) converges to f and (g_n) converges to g , where f, g are nonidentity transformations that have a common fixed point in $\overline{\mathbb{R}}$. Then for all n large enough, $\langle f_n, g_n \rangle$ is not semidiscrete and inverse-free.*

PROOF. Let $S_n = \langle f_n, g_n \rangle$ and $S = \langle f, g \rangle$. Firstly, note that due to Lemma 5.21, we have that $|\mathrm{tr}([h, k])| = 2$ for all transformations h, k in S . Also, the commutator is a continuous function from $\mathrm{PSL}(2, \mathbb{R})^2$ to $\mathrm{PSL}(2, \mathbb{R})$, which implies that for any $h_n, k_n \in S_n$ with $h_n \rightarrow h$ and $k_n \rightarrow k$, the sequence $|\mathrm{tr}([h_n, k_n])|$ converges to 2. We are going to use this fact throughout the proof.

Let us assume that S_n is semidiscrete and inverse-free, for infinitely many n . By relabelling the sequence S_n , we can assume that S_n is semidiscrete and inverse-free for all n . Since the transformations f, g have a fixed point in $\overline{\mathbb{R}}$, and none of them is the identity, each of the f, g is either hyperbolic or parabolic. Suppose that f is parabolic. Note that the trace of a Möbius transformation is continuous in $\mathrm{PSL}(2, \mathbb{R})$, and so $\mathrm{tr}(f_n)$ converges to 2 as $n \rightarrow \infty$. So, since the trace of the commutator $[f_n, g_n]$ also converges to 2, we have that the quantity

$$||\mathrm{tr}(f_n)|^2 - 4| + ||\mathrm{tr}([f_n, g_n])| - 2|,$$

is less than one, for all n large enough. Hence, Jørgensen's inequality for semigroups, Theorem 5.19, implies that in order for S_n to be semidiscrete, it would have to be a Schottky semigroup of rank one. This is a contradiction because f_n and g_n are antiparallel, and so S_n cannot be semidiscrete. We reach the same conclusion if g is parabolic.

So for the rest of the proof we can assume that f and g are hyperbolic. Note that the set of hyperbolic transformations of $\mathrm{PSL}(2, \mathbb{R})$ is open. So because f_n and g_n converge to f and g , respectively, they have to be hyperbolic transformations, for all n large enough. Thus, by relabelling the sequences (f_n) and (g_n) , we can assume that they are sequences of hyperbolic transformations. Also, since f_n and g_n are antiparallel, we have that the

attracting fixed point of one of the f and g is the repelling fixed point of the other. Without loss of generality, we assume that $\beta(f) = \alpha(g)$, and conjugate by a Möbius transformation so that $\alpha(f) < \beta(f) = \alpha(g) < \beta(g)$ (see Figure 5.3 on the right).

Due to our conjugation, for all n large enough we have that $\alpha(f_n) < \beta(f_n) < \alpha(g_n) < \beta(g_n)$ (see Figure 5.3 on the left). Because $\beta(f) = \alpha(g)$, Lemma 4.8 is applicable to the semigroup S and implies that the closure of S in $\text{PSL}(2, \mathbb{R})$ contains parabolic transformations that fix $\beta(f) = \alpha(g)$ and map the interval $(\alpha(f), \alpha(g))$ strictly inside itself. So, there exists a sequence (ϕ_n) in S that converges to such a parabolic transformation. By the convergence of (f_n) and (g_n) to f and g , respectively, we can pass to a subsequence (S_{n_k}) of (S_n) in order to obtain a sequence of transformations (h_k) in S_{n_k} such that h_k converges to a parabolic transformation fixing $\beta(f) = \alpha(g)$ and mapping $(\alpha(f), \alpha(g))$ strictly inside itself. For convenience we relabel the sequences (h_k) and (S_{n_k}) so that h_n lies in S_n , for all n .

Note that because S_n is semidiscrete and inverse-free, h_n have to be either hyperbolic or parabolic, for all n .

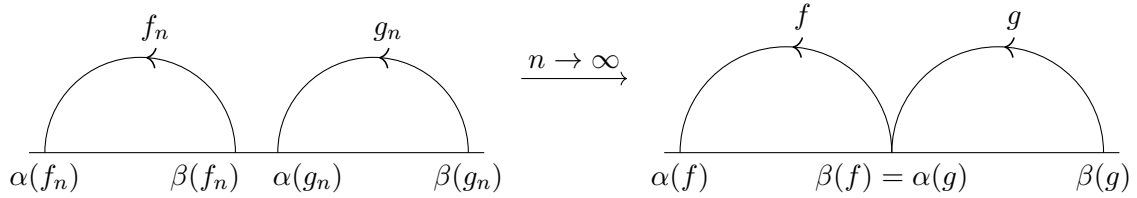


FIGURE 5.3. The configuration of f and g .

Suppose that h_n and one of f_n or g_n are antiparallel, for infinitely many n . By passing to a subsequence if necessary and without loss of generality, we can assume that h_n and f_n are antiparallel, for all $n \in \mathbb{N}$ (see the top part of Figure 5.4). Then Jørgensen's inequality for semigroups, Theorem 5.19, implies that

$$(5.3.1) \quad ||\text{tr}(h_n)|^2 - 4| + ||\text{tr}([f_n, h_n])| - 2| \geq 1,$$

for all positive integers n . But the left-hand side of (5.3.1) converges to zero, as h_n converges to a parabolic transformation and $\text{tr}([f_n, h_n])$ converges to 2, which is a contradiction. We reach the same contradiction by assuming that h_n and g_n are antiparallel, for all $n \in \mathbb{N}$.

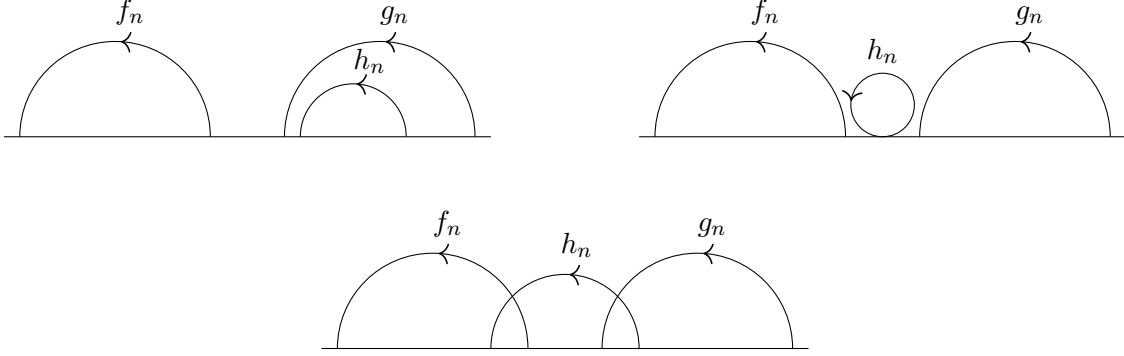


FIGURE 5.4. The sequence of transformations (h_n) .

Finally, we assume that none of the pairs h_n, f_n and h_n, g_n are antiparallel, for any n large enough. Observe that if h_n is parabolic, then one of the pairs h_n, f_n and h_n, g_n has to be antiparallel. Hence, all the transformations h_n have to be hyperbolic, for all n large enough. Also, due to the fact that the sequences of points $\beta(f_n), \alpha(h_n), \beta(h_n)$ and $\alpha(g_n)$ all converge to the same point $\beta(f) = \alpha(g)$, the axis of h_n has to intersect both $\text{Ax}(f_n)$ and $\text{Ax}(g_n)$, for all n large enough (see the bottom part of Figure 5.4). As the sequence h_n converges to a parabolic transformation that fixes $\beta(f) = \alpha(g)$ and maps $(\alpha(f), \beta(f))$ strictly inside itself, we must have that $\alpha(h_n) < \beta(f_n)$, for all n large enough. Thus, we have the following order for the fixed points of the hyperbolic transformations f_n, h_n and g_n

$$(5.3.2) \quad \alpha(f_n) < \alpha(h_n) < \beta(f_n) < \alpha(g_n) < \beta(h_n) < \beta(g_n), \quad \text{for all } n \text{ large enough.}$$

Note that the sequence $(h_n(\alpha(f_n)))$ converges some point in the open interval $(\alpha(f), \beta(f))$, and the sequence $(f_n(\alpha(h_n)))$ converges to $\beta(f)$. Hence, $h_n(\alpha(f_n)) \leq f_n(\alpha(h_n))$, for all n large enough. Then we can apply Lemma 4.20 to the transformations f_n, h_n in order to obtain that $\Lambda^+(\langle f_n, h_n \rangle) = [\alpha(f_n), \alpha(h_n)]$. Furthermore, by the convergence of $(f_n), (g_n)$ and (h_n) , the point $\beta(f_n)$ is contained in the interior of the interval $g_n([\alpha(f_n), \alpha(h_n)]) \subset$

$\Lambda^+(S_n)$, for all n large enough. Therefore, the points $(\beta(f_n))$ lie in the interior of $\Lambda^+(S_n)$ for all n large enough. Lemma 4.12 is then applicable to the semigroup S_n and implies that it is either a discrete group or it is not semidiscrete. But, S_n is inverse-free, so it cannot be a group and we have reached a contradiction.

All our contradictions were reached because we assumed that there exists a subsequence of S_n that is semidiscrete and inverse-free. Therefore, we conclude that the semigroups S_n are not semidiscrete and inverse-free for all large enough n . \square

THEOREM 5.23. *Let (\mathcal{F}_m) be a sequence in \mathcal{S} that converges to a point $\mathcal{F} \in \mathcal{S}$. If $\langle \mathcal{F} \rangle$ is a Schottky semigroup of rank one, then for all m large enough $\langle \mathcal{F}_m \rangle$ is also a Schottky semigroup of rank one.*

PROOF. Write $\mathcal{F} = (f_1, f_2, \dots, f_N)$ and $\mathcal{F}_m = (f_1^m, f_2^m, \dots, f_N^m)$. Also, without loss of generality, suppose that $(0, 1) \subset \overline{\mathbb{R}}$ is mapped strictly inside itself by $\langle \mathcal{F} \rangle$. Since $\langle \mathcal{F} \rangle$ is semidiscrete and inverse-free and maps an interval strictly inside itself, all ordinates of \mathcal{F} have to either be hyperbolic transformations with attracting fixed points in $[0, 1]$ and repelling fixed points in the complement of $(0, 1)$, or parabolic transformations that fix 0 or 1 and map $(0, 1)$ inside itself. Let us assume that there exists a subsequence of (\mathcal{F}_m) , such that $\langle \mathcal{F}_{m_k} \rangle$ is semidiscrete and inverse-free but not a Schottky semigroup of rank one, for all k . By relabelling the sequence (\mathcal{F}_{m_k}) , we can assume that $\langle \mathcal{F}_m \rangle$ is semidiscrete and inverse-free but not a Schottky semigroup of rank one, for all m .

We are first going to show that \mathcal{F}_m has to contain pairs of ordinates that are antiparallel, for all m large enough. Let us assume otherwise. Then in order for $\langle \mathcal{F}_m \rangle$ to not be a Schottky semigroup of rank one, for all large enough m , there must exist ordinates f_i^m, f_j^m and f_l^m of \mathcal{F}_m with the following properties, up to conjugation: f_i^m and f_j^m are hyperbolic transformations, with crossing axes and such that $\alpha(f_i^m) < \alpha(f_j^m) < \beta(f_i^m) < \beta(f_j^m)$; the axis of f_l^m intersects $\text{Ax}(f_i^m)$ and $\text{Ax}(f_j^m)$, and the repelling fixed point of f_l^m (or its unique fixed point if it is parabolic) lies in the interval $(\alpha(f_i^m), \alpha(f_j^m))$ (see Figure 5.5). Also, by passing to a further subsequence if necessary, we can assume that f_i^m, f_j^m and f_l^m converge to some ordinates f_i, f_j and f_l of \mathcal{F} , respectively. Note that the attracting fixed points of f_i^m, f_j^m and f_l^m converge to some points in $[0, 1]$. In addition, the transformations f_i, f_j and f_l have to map the interval $(0, 1)$ strictly inside itself. Hence, the points $\alpha(f_i^m), \beta(f_l^m)$

and $\alpha(f_j^m)$ all converge to some endpoint of $(0, 1)$, say 0. In addition, because the axis of f_l^m intersects $\text{Ax}(f_i^m)$ and $\alpha(f_i^m) \leq \beta(f_l^m)$, we have that $\beta(f_i^m) \leq \alpha(f_l^m)$ for all m . But $\alpha(f_l^m)$ converges to some point in $[0, 1]$, and so the transformations f_i^m and f_l^m have the same axes and $\alpha(f_i^m) = \beta(f_l^m)$. This implies that $\langle \mathcal{F}_m \rangle$ is not semidiscrete and inverse-free and we have reached a contradiction.

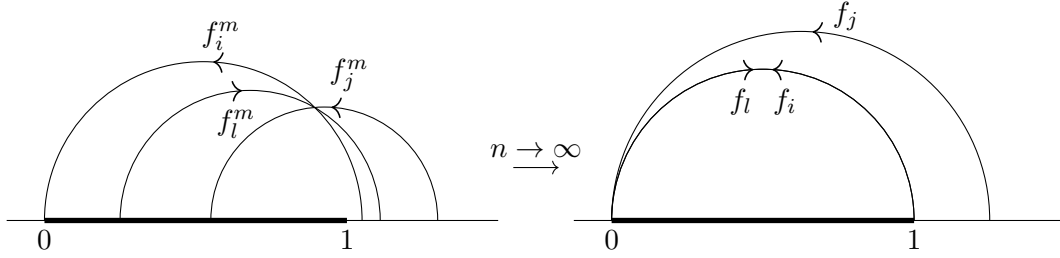
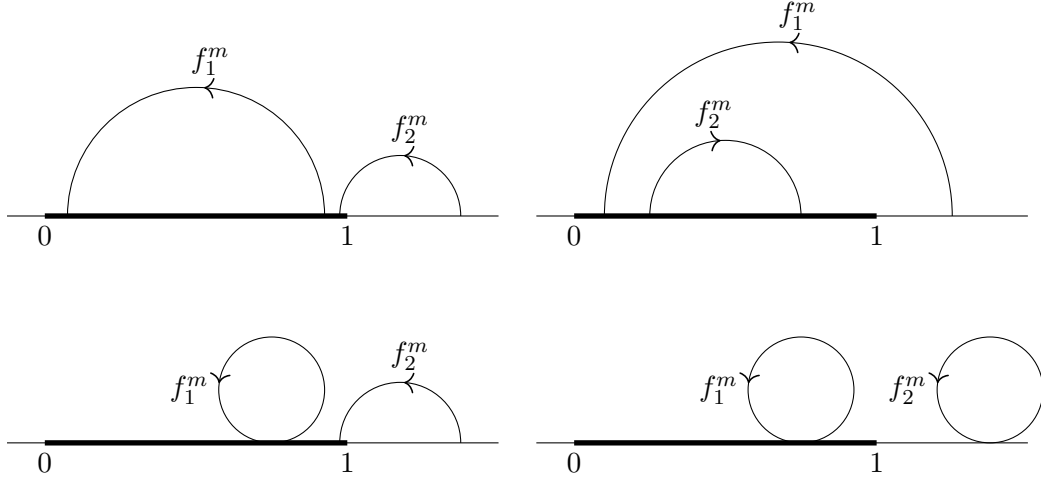


FIGURE 5.5. $\langle \mathcal{F}_m \rangle$ does not contain any antiparallel pairs.

So the fact that $\langle \mathcal{F}_m \rangle$ is not a Schottky semigroup of rank one implies that we can find ordinates f_i^m and f_j^m of \mathcal{F}_m that are antiparallel, for all m large enough. Without loss of generality, assume that f_1^m and f_2^m are antiparallel for all m (see Figure 5.6). By passing to a subsequence if necessary, we can assume that f_1^m and f_2^m converge to the ordinates f_1 and f_2 of \mathcal{F} , respectively. Because the attracting fixed points (or the unique fixed points if they are parabolic) of f_1 and f_2 have to lie in $[0, 1]$, and the pair f_1, f_2 is approximated by a pair of antiparallel transformations, we have that f_1 and f_2 both fix an endpoint of $(0, 1)$, say 1. Then Lemma 5.22 is applicable to the sequences (f_1^m) and (f_2^m) , and yields that $\langle \mathcal{F}_m \rangle$ is not semidiscrete and inverse-free, for any m large enough, which is a contradiction.

We conclude that for all n large enough, \mathcal{F}_n generates a Schottky semigroup of rank one. \square

Observe that Theorem 5.23 and Lemma 5.16 imply that the principle component is isolated in the parameter space \mathcal{H} , in the sense that $\overline{H_P}$ is disjoint from the closure of all non-principle components. Theorem 5.17, however, implies that the principal component is also isolated from all other components of the parameter space \mathcal{S} , which is a significantly

FIGURE 5.6. $\langle \mathcal{F}_m \rangle$ contains antiparallel pairs.

stronger statement.

We are now ready to prove Theorem 5.17.

PROOF OF THEOREM 5.17. Suppose that (\mathcal{F}_m) is a sequence in $\mathcal{S} \setminus \overline{H_P}$ that converges to some $\mathcal{F} \in \text{PSL}(2, \mathbb{R})^N$. We need to prove that \mathcal{F} lies in $\mathcal{S} \setminus \overline{H_P}$. Recall that Lemma 5.16 implies that N -tuples that generate Schottky semigroups of rank one lie in the closure of the principal component. Therefore the semigroups $\langle \mathcal{F}_m \rangle$ cannot be Schottky semigroups of rank one. So due to Theorem 5.23, \mathcal{F} cannot lie in the closure of the principal component, and we only need to show that it lies in \mathcal{S} ; that is, $\langle \mathcal{F} \rangle$ is semidiscrete and inverse-free. Assume, towards a contradiction, that there exists a sequence of transformations in $\langle \mathcal{F} \rangle$ that converges to the identity. Then, from the convergence of \mathcal{F}_m to \mathcal{F} , we can find $f_m \in \langle \mathcal{F}_m \rangle$, for all $m \geq 0$, such that (f_m) converges to the identity. By passing on to a subsequence if necessary, we can assume that $\alpha(f_m)$ and $\beta(f_m)$ converge to some points α and β in $\overline{\mathbb{R}}$.

Suppose that for all m there exist ordinates g_m of \mathcal{F}_m such that f_m and g_m are antiparallel. We may assume that (g_m) converges to some ordinate g of \mathcal{F} . Then Jørgensen's inequality for semigroups, Theorem 5.19, implies that

$$(5.3.3) \quad ||\text{tr}(f_m)|^2 - 4| + ||\text{tr}([f_m, g_m])| - 2| \geq 1.$$

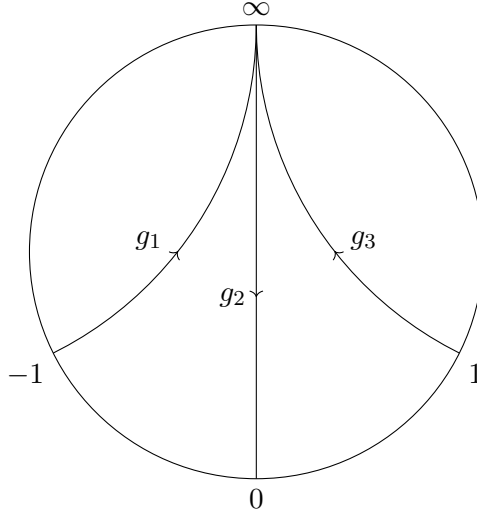
But because f_m converges to the identity and g_m to g , the commutator $[f_m, g_m] = f_m g_m f_m^{-1} g_m^{-1}$ converges to the identity. So the left-hand side of inequality (5.26) converges to zero, as $m \rightarrow \infty$, which is a contradiction.

Assume now that there are no antiparallel pairs of ordinates of \mathcal{F}_m , for any m . Then since $\langle \mathcal{F}_m \rangle$ is not a Schottky semigroup of rank one, for every $m \in \mathbb{N}$ we can find an ordinate h_m of \mathcal{F}_m , such that f_m and h_m map an open interval I_m inside itself and $\Lambda^-(\mathcal{F}_m) \cap I_m \neq \emptyset$. After a suitable conjugation, we can assume that h_m and I_m have the following properties: $h_m(x_m) = x_m$, for some point x_m in $\overline{\mathbb{R}}$ with $x_m < \alpha(f_m)$, and $I_m = (x_m, \alpha(f_m))$. We can also assume that (h_m) converges to some ordinate h of \mathcal{F} , and x_m converges to some $x \leq \alpha$. Then, since (f_m) converges to the identity and $\tau(h_m)$ is bounded above, we have that $f_m(x_m) \leq h_m(\alpha(f_m))$, for all m large enough. Hence, we can apply Lemma 4.20 to the transformations f_n and h_n in order to obtain that $\Lambda^+(\langle f_m, h_m \rangle) = [x_m, \alpha(f_m)] = \overline{I_m}$, for all m large enough. But this implies that $\emptyset \neq \Lambda^-(\mathcal{F}_m) \cap I_m \subseteq \Lambda^-(\mathcal{F}_m) \cap \Lambda^+(\mathcal{F}_m)^\circ$. Lemma 4.12 is then applicable to the semigroup $\langle \mathcal{F}_m \rangle$ and yields that either it is a discrete group or it is not semidiscrete, which is a contradiction.

In conclusion, contrary to our assumption, $\langle \mathcal{F} \rangle$ is semidiscrete and inverse-free. \square

We are now ready to present our example that answers Question 5.11. We start by presenting an example for the case $N = 3$, which we will then extend to any $N \geq 3$.

EXAMPLE 5.24. Suppose that $N = 3$ and consider the following hyperbolic transformations: $g_1(z) = 2z + 1$, $g_2(z) = \frac{1}{3}z$ and $g_3(z) = 5z - 4$ (see Figure 5.7 for the axes of the transformations g_1, g_2 and g_3). Observe that the semigroup generated by the 3-tuple $\mathcal{F}_0 = (g_1, g_2, g_3)$ is elementary, since g_1, g_2 and g_3 all fix the point infinity. Also, according to Lemma 4.9 it is not semidiscrete. It is easy to check that any element in $\langle \mathcal{F}_0 \rangle$ is of the form $\lambda z + \kappa$, where $\lambda = 2^l 3^{-m} 5^n$, for some $l, m, n \in \mathbb{N} \cup \{0\}$, and some real number κ . Note that l, m and n cannot all be 0 simultaneously, and so because 2, 3 and 5 are prime numbers, λ cannot be 1. Thus all the transformations in $\langle \mathcal{F}_0 \rangle$ are hyperbolic, which implies that the semigroup $\langle \mathcal{F}_0 \rangle$ does not contain any elliptic transformations or the identity. Hence, the 3-tuple \mathcal{F}_0 lies in the complement of \mathcal{E}_I in $\text{PSL}(2, \mathbb{R})^3$.

FIGURE 5.7. The 3-tuple $\mathcal{F}_0 = (g_1, g_2, g_3)$

In order to show that Example 5.24 answers Question 5.11 in the negative, we need to prove that \mathcal{F}_0 lies in the complement of $\overline{\mathcal{H}}$. Note that because $\mathcal{H} \subset \mathcal{S}$, it suffices to prove that \mathcal{F}_0 lies in the complement of $\overline{\mathcal{S}}$.

Recall that the set $\mathcal{S} \setminus \overline{H_P}$ is closed, as stated by Theorem 5.17. Therefore, because $\langle \mathcal{F}_0 \rangle$ is not semidiscrete, if \mathcal{F}_0 were to lie in $\overline{\mathcal{S}}$, it would have to lie in the closure of the principal component. The following lemma tells us that this cannot happen, thus proving our claim.

LEMMA 5.25. *The 3-tuple \mathcal{F}_0 in Example 5.24 lies in the complement of $\overline{H_P}$.*

PROOF. It suffices to show that there exists an open neighbourhood of \mathcal{F}_0 in $\text{PSL}(2, \mathbb{R})^3$ that is disjoint from the principal component H_P . Take $\varepsilon > 0$ and define the open intervals $U_{-1} = (-1 - \varepsilon, -1 + \varepsilon)$, $U_0 = (-\varepsilon, \varepsilon)$ and $U_1 = (1 - \varepsilon, 1 + \varepsilon)$. Also define $U_\infty = [\infty, -1/\varepsilon) \cup (1/\varepsilon, \infty]$, which is a neighbourhood of ∞ in $\overline{\mathbb{R}}$. We can choose ε small enough such that the sets U_{-1}, U_0, U_1 and U_∞ have pairwise disjoint closures. Note that the set of hyperbolic transformations in $\text{PSL}(2, \mathbb{R})$ is open. So there exists an open neighbourhood D of \mathcal{F}_0 , such that if (f_1, f_2, f_3) lies in D , then f_1, f_2 and f_3 are hyperbolic transformations. We can also choose D small enough so that for all $(f_1, f_2, f_3) \in D$ we have $\beta(f_1) \in U_{-1}$, $\alpha(f_2) \in U_0$, $\beta(f_3) \in U_1$ and $\alpha(f_1), \beta(f_2), \alpha(f_3) \in U_\infty$ (see Figure 5.8). Hence, by the configuration of the axes of f_1, f_2 and f_3 we see that (f_1, f_2, f_3) cannot

generate a Schottky semigroup of rank one, and this implies that D does not intersect H_P , as required. \square

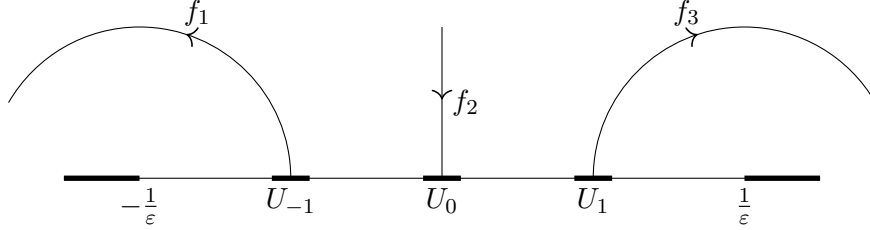


FIGURE 5.8. An open neighbourhood U of \mathcal{F}_0

We now extend Example 5.24 to any $N \geq 3$.

DEFINITION 5.26. We define \mathcal{P} to be the set of N -tuples that generate elementary and inverse-free semigroups, that are not semidiscrete and do not contain elliptic transformations.

First, let us establish that \mathcal{P} is non-empty for all $N \geq 3$. If $N = 3$, then the 3-tuple \mathcal{F}_0 in Example 5.24 lies in \mathcal{P} . Suppose that $N > 3$. Consider the transformations $g_1(z) = 2z + 1$, $g_2(z) = \frac{1}{3}z$ and $g_3(z) = 5z - 4$ that were introduced in Example 5.24. Let $\mathcal{F} = (g_1, g_2, g_3, g_4, \dots, g_N)$, where g_i is a hyperbolic transformation that fixes the point infinity and does not fix any of the points $-1, 0$ and 1 , for all $i = 4, 5, \dots, N$. Hence, for all $i = 4, 5, \dots, N$, g_i is of the form $g_i(z) = \lambda_i z + \kappa_i$, for some $\lambda_i > 0$ and $\kappa_i \in \mathbb{R}$. We also choose λ_i to be either p or $\frac{1}{p}$, where p is a prime number different from $2, 3$ and 5 . So it is easy to see that all elements in $\langle \mathcal{F} \rangle$ are hyperbolic transformations that fix the point infinity, and thus $\langle \mathcal{F} \rangle$ is elementary and inverse-free. In addition, because $\langle \mathcal{F}_0 \rangle \subset \langle \mathcal{F} \rangle$, the semigroup $\langle \mathcal{F} \rangle$ is not semidiscrete. We conclude that the N -tuple \mathcal{F} lies in \mathcal{P} .

Observe that \mathcal{P} does not intersect the elliptic locus. We finish this chapter by proving the following description for the complement of the semidiscrete and inverse-free locus. Note that because $\mathcal{H} \subset \mathcal{S}$, Theorem 5.27 answers Question 5.11 in the negative for any $N \geq 3$.

THEOREM 5.27. For all $N \geq 3$, we have that $(\mathcal{S} \setminus \overline{H_P})^c = \mathcal{E}_I \cup \overline{H_P} \cup \mathcal{P}$ in $\text{PSL}(2, \mathbb{R})^N$.

Observe that $\mathcal{E}_I \cup \overline{H_P} \cup \mathcal{P}$ is not a disjoint union (for example $\mathcal{E}_I \cap \overline{H_P} \neq \emptyset$), but due to Theorem 5.17 it is an open subset of $\mathrm{PSL}(2, \mathbb{R})^N$. In order to prove Theorem 5.27 we require the following classification of finitely-generated semigroups due to Jacques and Short [26, Theorem 14.1] (recall Definition 5.15).

THEOREM 5.28 ([26]). *Let S be a finitely generated semigroup. Then S is*

- (1) *elementary;*
- (2) *semidiscrete;*
- (3) *contained in $\mathcal{M}(J)$, for some J ; or*
- (4) *dense in $\mathrm{PSL}(2, \mathbb{R})$.*

We remark that the four classes of semigroups in Theorem 5.28 are not all disjoint. For example, there exist elementary and semidiscrete semigroups that lie in $\mathcal{M}(J)$, for some J . However, semidiscrete semigroups cannot be dense in $\mathrm{PSL}(2, \mathbb{R})$ and vice versa.

PROOF OF THEOREM 5.27. The inclusion $\mathcal{E}_I \cup \overline{H_P} \cup \mathcal{P} \subseteq (\mathcal{S} \setminus \overline{H_P})^c$ is obvious from the definitions of \mathcal{E}_I , H_P and \mathcal{P} . For the converse, let \mathcal{F} be an N -tuple in $(\mathcal{S} \setminus \overline{H_P})^c$, and suppose that $\mathcal{F} \notin (\mathcal{E}_I \cup \overline{H_P})$. Our task is to show that \mathcal{F} lies in \mathcal{P} . Note that because $\mathcal{F} \notin \mathcal{E}_I$, the semigroup $\langle \mathcal{F} \rangle$ is inverse-free and does not contain elliptic transformations. Corollary 5.18 implies that $(\mathcal{S} \cup \overline{H_P})^c = \overline{\mathcal{S}}^c$, and so the semigroup generated by \mathcal{F} is not semidiscrete. Hence, we only need to show that $\langle \mathcal{F} \rangle$ is elementary.

Observe that due to Lemma 5.16, the semigroup $\langle \mathcal{F} \rangle$ cannot be contained in $\mathcal{M}(J)$ because then it would have to lie in the closure of the principal component. Also, since the set of elliptic transformations of $\mathrm{PSL}(2, \mathbb{R})$ is open, $\langle \mathcal{F} \rangle$ cannot be dense in $\mathrm{PSL}(2, \mathbb{R})$ as then \mathcal{F} would lie in \mathcal{E}_I . Finally, from Theorem 5.28, $\langle \mathcal{F} \rangle$ has to be elementary and thus \mathcal{F} lies in \mathcal{P} , as required. \square

CHAPTER 6

Boundaries of the hyperbolic components

6.1. Outline of the chapter

In this chapter we investigate the structure of the non-principal components of the hyperbolic locus, and in particular, the properties of semigroups generated by N -tuples that lie on the boundary of a non-principal component. Our main goal is to answer the following question of Avila, Bochi and Yoccoz [2, Question 2].

QUESTION 6.1 ([2]). Is the union of the boundaries of the components of \mathcal{H} in $\mathrm{PSL}(2, \mathbb{R})^N$ equal to the boundary of \mathcal{H} ?

We are going to answer Question 6.1 in the negative by constructing an N -tuple that lies on the boundary of \mathcal{H} but not on the boundary of any hyperbolic component. This construction will be carried out in Section 6.3.

In the next section we prove that N -tuples on the boundary of non-principal components generate Schottky semigroups.

6.2. Schottky semigroups

We now study N -tuples that generate Schottky semigroups. Recall that due to Lemma 5.16, all Schottky semigroups of rank one lie in the closure of the principal component H_P . The next result can be thought of as an extension of [2, Lemma 4.12] to certain N -tuples in \mathcal{S} , and in fact the proofs of these two lemmas are similar.

LEMMA 6.2. *Suppose that \mathcal{F} lies in $\mathcal{S} \setminus \overline{H_P}$ and $\langle \mathcal{F} \rangle$ is a Schottky semigroup with $\Lambda^+(\mathcal{F}) \cap \Lambda^-(\mathcal{F}) \neq \emptyset$. Then for each point $x \in \Lambda^+(\mathcal{F}) \cap \Lambda^-(\mathcal{F})$ there exist points $a, b \in \Lambda^+(\mathcal{F}) \cap \Lambda^-(\mathcal{F})$ and transformations $f, g \in \langle \mathcal{F} \rangle \cup \{\mathrm{Id}\}$ such that $x = f(a) = g^{-1}(b)$. In addition, the points a and b are fixed points of transformations in $\langle \mathcal{F} \rangle$.*

PROOF. Because $\langle \mathcal{F} \rangle$ is a Schottky semigroup, Lemma 5.5 implies that $C^+(\mathcal{F})$ and $C^-(\mathcal{F})$ are finitely connected and $\Lambda^+(\mathcal{F}) \cap \Lambda^-(\mathcal{F}) = \partial C^+(\mathcal{F}) \cap \partial C^-(\mathcal{F})$. Thus the intersection $\Lambda^+(\mathcal{F}) \cap \Lambda^-(\mathcal{F})$ is finite.

Let $x \in \Lambda^+(\mathcal{F}) \cap \Lambda^-(\mathcal{F})$. Since x is in the forward limit set of $\langle \mathcal{F} \rangle$, there exists a sequence (k_n) in $\langle \mathcal{F} \rangle$ such that $k_n(i) \rightarrow x$, as $n \rightarrow \infty$. Using Lemma 5.3, and by passing to a subsequence if necessary, we can write k_n as $k_n = f_i h_n$, for some ordinate f_i of \mathcal{F} and some sequence (h_n) in $\langle \mathcal{F} \rangle$. Then $h_n = f_i^{-1} k_n$, for all n , and so $(h_n(i))$ converges to $f_i^{-1}(x) \in \overline{\mathbb{R}}$. Thus, because $h_n \in \langle \mathcal{F} \rangle$, for all n , the point $f_i^{-1}(x)$ lies in the forward limit set of $\langle \mathcal{F} \rangle$. But x also lies in $\Lambda^-(\mathcal{F})$, which is a backward invariant set, and so $f_i^{-1}(x)$ also lies in the backward limit set of $\langle \mathcal{F} \rangle$.

Hence, for every $x \in \Lambda^+(\mathcal{F}) \cap \Lambda^-(\mathcal{F})$ there exists an ordinate f_i of \mathcal{F} such that $f_i^{-1}(x) \in \Lambda^+(\mathcal{F}) \cap \Lambda^-(\mathcal{F})$. Arguing similarly, we can show that there exists an ordinate f_j of \mathcal{F} such that $f_j(x) \in \Lambda^+(\mathcal{F}) \cap \Lambda^-(\mathcal{F})$. Iterating these processes and using the finiteness of $\Lambda^+(\mathcal{F}) \cap \Lambda^-(\mathcal{F})$, we find points $a, b \in \Lambda^+(\mathcal{F}) \cap \Lambda^-(\mathcal{F})$ which are fixed points of transformations in $\langle \mathcal{F} \rangle$ and transformations $f, g \in \langle \mathcal{F} \rangle$, such that $f^{-1}(x) = a$ and $g(x) = b$, which yields the desired conclusion. \square

We finish this section with an extension of [2, Proposition 4.9], which states that if \mathcal{F} lies in the closure of a non-principal component, then $\langle \mathcal{F} \rangle$ is inverse-free.

THEOREM 6.3. *Suppose that \mathcal{F} lies on the boundary of a non-principal component H of \mathcal{H} . Then $\langle \mathcal{F} \rangle$ is a Schottky semigroup of rank greater than one.*

PROOF. Observe that \mathcal{F} lies in $\partial H \subset \mathcal{S} \setminus \overline{H_P}$, and thus $\langle \mathcal{F} \rangle$ is semidiscrete and inverse-free due to Theorem 5.17. Also, because \mathcal{F} does not lie in the closure of the principal component, if $\langle \mathcal{F} \rangle$ were a Schottky semigroup, its rank would have to be greater than one. Suppose that $\mathcal{F}_n \in H$ is a sequence of N -tuples with $\mathcal{F}_n \rightarrow \mathcal{F}$, in the topology of $\text{PSL}(2, \mathbb{R})^N$, and define $C_n^+ = C^+(\mathcal{F}_n)$ and $C_n^- = C^-(\mathcal{F}_n)$. Under these assumptions, [2, Proposition 4.13] states that C_n^+ and C_n^- converge, in the Hausdorff metric, to some sets C^+ and C^- that, by continuity, are finitely connected, nonempty subsets of $\overline{\mathbb{R}}$. We are going to prove that C^+ is mapped strictly inside itself by each ordinate of \mathcal{F} , and it is a union of nontrivial closed intervals, none of which are singletons.

First, take $x \in C^+$ and let f be an ordinate of \mathcal{F} . There exist sequences of points (x_n) in C_n^+ and ordinates (f_n) of \mathcal{F}_n , such that $x_n \rightarrow x$ and $f_n \rightarrow f$, as $n \rightarrow \infty$. Conjugating by a Möbius transformation, we can assume that $x = 1$ and $f(z) = \lambda z$, for some positive $\lambda \in \mathbb{R}$. Also, without loss of generality, we assume that $x_n \leq 1$, for all n . Applying the mean value theorem to the interval $(x_n, 1)$, for some n , we obtain that there exists $\xi_n \in (x_n, 1)$, such that

$$|f_n(1) - f_n(x_n)| = |f'_n(\xi_n)| |1 - x_n|.$$

But $f'_n(\xi_n)$ converges to λ , and so $f_n(x_n)$ converges to the point $f(1)$, as $n \rightarrow \infty$. Since \mathcal{F}_n lie in a non-principal hyperbolic component, $\langle \mathcal{F}_n \rangle$ is Schottky semigroup of rank greater than one. Hence, Lemma 5.5 implies that C_n^+ is forward invariant under $\langle \mathcal{F}_n \rangle$. So the points $f_n(x_n)$ lie in C_n^+ , for all n . Thus, by the convergence of C_n^+ to C^+ , we have that $f(1) \in C^+$. We conclude that C^+ is mapped inside itself by each element of $\langle \mathcal{F} \rangle$.

Assume now that C^+ has a connected component $\{x\}$, for some $x \in \overline{\mathbb{R}}$. The second part of Lemma 5.5 implies that x lies in $\Lambda^+(\mathcal{F})$. As $\Lambda^+(\mathcal{F})$ is the closure of all attracting fixed points of $\langle \mathcal{F} \rangle$, we must have that $x = \alpha(h)$, for some hyperbolic transformation $h \in \langle \mathcal{F} \rangle$. But C^+ is mapped inside itself by $\langle \mathcal{F} \rangle$ and so $C^+ = \{x\}$. Thus \mathcal{F} has to lie in $\overline{H_P}$, which is a contradiction.

Using similar arguments we can prove that C^- is mapped inside itself by each element of $\langle \mathcal{F} \rangle^{-1}$, and does not contain any connected components that are singletons. Hence, the set C^+ cannot be the extended real line, because otherwise the convergence of C_n^+ and C_n^- would imply that C^- is a finite subset of $\overline{\mathbb{R}}$.

We will now prove that $\langle \mathcal{F} \rangle$ maps C^+ strictly inside itself. Suppose that there exists an ordinate g of \mathcal{F} that fixes C^+ . Since C^+ cannot be empty, the extended real line or have components that are singletons, C^+ has to be an interval. Then $\langle \mathcal{F} \rangle$ lies in $\mathcal{M}(C^+)$, which by Lemma 5.16 implies that \mathcal{F} lies in the closure of the principal component, a contradiction.

We conclude that C^+ is a union of closed, disjoint intervals that is mapped strictly inside itself by $\langle \mathcal{F} \rangle$, and thus $\langle \mathcal{F} \rangle$ is a Schottky semigroup. \square

6.3. Semigroups within groups

We now construct a counterexample to Question 6.1. In particular, we find hyperbolic transformations f and g that generate a discrete group, and such that the 4-tuple

$$(f, g, f^{-1}gf, fg^{-1}f^{-1})$$

lies on the boundary of \mathcal{H} but does not lie on the boundary of any component of \mathcal{H} . This also acts as a more elaborate example of the fact that the converse of Theorem 5.13 is not true in general.

First, as promised, let us formally define Schottky groups.

DEFINITION 6.4. Let $\{f_1, f_2, \dots, f_N\}$ be a collection in $\mathrm{PSL}(2, \mathbb{R})$. Suppose that for all $k = 1, 2, \dots, N$ there exist collections of open intervals A_k and B_k in $\overline{\mathbb{R}}$, with disjoint closures, such that f_k maps the complement of $\overline{B_k}$ onto A_k , for all $k = 1, 2, \dots, N$. Then, the group generated by $\{f_1, f_2, \dots, f_N\}$ is called a *Schottky group*.

Schottky groups as we have defined them are often called *classical Schottky groups* in the literature. It is well-known that every Schottky group is a discrete, free and purely hyperbolic group (see, for example, [34, Section 2.7], [35] and [36, Section X.H.]).

EXAMPLE 6.5. Let

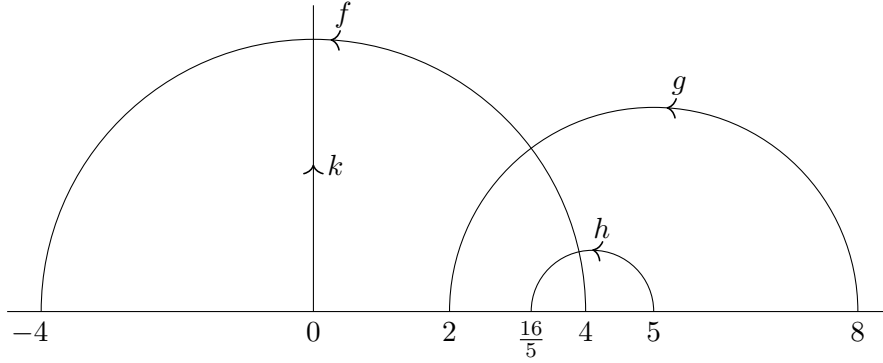
$$f(z) = \frac{-8z + 16}{z - 8}.$$

This is a hyperbolic Möbius transformation with attracting fixed point -4 and repelling fixed point 4 . Let $k(z) = \lambda z$, where $\lambda > 40$.

Define $g = f^{-1}k^{-1}f$. Then $k = fg^{-1}f^{-1}$. The attracting and repelling fixed points of g are $f^{-1}(0) = 2$ and $f^{-1}(\infty) = 8$, respectively.

Last, define $h = f^{-2}k^{-1}f^2$. Then $h = f^{-1}gf$. The attracting and repelling fixed points of h are $f^{-2}(0) = 16/5$ and $f^{-2}(\infty) = 5$, respectively.

The transformation f maps $\overline{\mathbb{R}} \setminus [8 - 4\sqrt{3}, 8 + 4\sqrt{3}]$ into $[-8 - 4\sqrt{3}, -8 + 4\sqrt{3}]$. And k maps $\overline{\mathbb{R}} \setminus [-1, 1]$ into $\overline{\mathbb{R}} \setminus [-15, 15]$. Since the four intervals $[8 - 4\sqrt{3}, 8 + 4\sqrt{3}]$, $[-8 - 4\sqrt{3}, -8 + 4\sqrt{3}]$, $[-1, 1]$ and $\overline{\mathbb{R}} \setminus [-15, 15]$ are pairwise disjoint, we see that f and k generate a Schottky group.

FIGURE 6.1. The transformations f, g, h and k .

Let us establish that (f, g, h, k) lies in \mathcal{S} . Since the semigroup $\langle f, g, h, k \rangle$ lies in a discrete group it is semidiscrete. Suppose that $\phi = \text{Id}$, for some $\phi \in \langle f, g, h, k \rangle$. The transformation ϕ can be written as a composition of positive powers of $f, g, h = f^{-1}gf$ and $k = fg^{-1}f^{-1}$. But $\langle f, g, h, k \rangle$ lies in a free group, so the powers of f and g in the finite word ϕ have to add up to zero, which is impossible.

In addition, (f, g, h, k) does not lie in the closure of the principal component since $\langle f, g, h, k \rangle$ is semidiscrete and inverse-free, but not a Schottky semigroup of rank one. If (f, g, h, k) were to lie on the boundary of a non-principal component of \mathcal{H} , then $\langle f, g, h, k \rangle$ would have to be a Schottky semigroup by Theorem 6.3. Observe, however, that $fg^n(\alpha(h)) = fg^n(16/5)$ accumulates at 0 on the right, and so does the sequence $k^{-n}(\beta(f)) = k^{-n}(4)$ (see Figure 6.1). Also, the sequences $fg^n(\alpha(h))$ and $k^{-n}(\beta(f))$ lie in the forward and backward limit sets of $\langle f, g, h, k \rangle$, respectively. Suppose that M is an open proper subset of \overline{R} that is mapped inside itself by $\langle f, g, h, k \rangle$. Then $fg^n(\alpha(h))$ lies in \overline{M} and $k^{-n}(\beta(f))$ lies in the complement of M . Hence, M is infinitely connected and thus $\langle f, g, h, k \rangle$ cannot be a Schottky semigroup. This also implies that (f, g, h, k) is not uniformly hyperbolic, since $0 \in \Lambda^+((f, g, h, k)) \cap \Lambda^-((f, g, h, k))$ (see Theorem 5.6).

In order to show that $(f, g, h, k) \in \overline{\mathcal{H}}$, we construct a sequence of 4-tuples (f, g, h, k_m) , $m = 1, 2, \dots$, in the hyperbolic locus that converges to (f, g, h, k) .

For $m = 0, 1, 2, \dots$, let

$$t_m(z) = z + \frac{2}{\lambda_{m+1}} \quad \text{and} \quad k_m = t_m k t_m^{-1}.$$

The attracting fixed point of k_m is ∞ and the repelling fixed point of k_m is $t_m(0) = 2/\lambda^{m+1}$.

Fix a positive integer m . For $n = 0, 1, \dots, m$, define

$$I_n = \left[k^n \left(\frac{1}{\lambda^m} \right), k^n \left(\frac{2}{\lambda^m} \right) \right] = \left[\frac{1}{\lambda^{m-n}}, \frac{2}{\lambda^{m-n}} \right],$$

$$J_n = \left[k_m^n \left(\frac{1}{\lambda^m} \right), k_m^n \left(\frac{2}{\lambda^m} \right) \right] = \left[\frac{2}{\lambda^{m+1}} + \frac{1-2/\lambda}{\lambda^{m-n}}, \frac{2}{\lambda^{m-n}} \right].$$

Observe that $I_n \subset J_n$, for $n = 0, 1, \dots, m$. Also, the intervals J_0, J_1, \dots, J_m are disjoint, and they are listed in increasing order.

Let

$$A = \overline{\mathbb{R}} - (2/\lambda^{m+1}, 10),$$

$$B = J_0 \cup J_1 \cup \dots \cup J_m,$$

$$C = f^{-1}([0, 2/\lambda^{m+1}]),$$

$$D = f^{-1}(I_0) \cup f^{-1}(I_1) \cup \dots \cup f^{-1}(I_m).$$

Define

$$X = A \cup B \cup C \cup D.$$

We will prove that f, g, h and k_m map X inside itself.

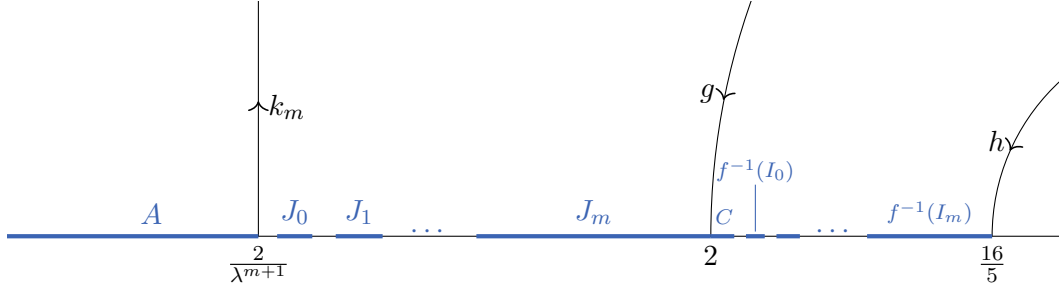


FIGURE 6.2. The action of $\langle f, g, h, k_m \rangle$ in $\overline{\mathbb{R}}$.

First we prove that $f(X) \subset X$. Observe that A is a closed interval in $\overline{\mathbb{R}}$ containing the attracting but not the repelling fixed point of f , so $f(A) \subset A$. Furthermore, $f(D) \subset B$, $f(B) \subset A$ and $f(C) \subset A$. Hence $f(X) \subset X$.

Next we prove that $g(X) \subset X$. Let $E = J_m \cup C$. The interval E contains the attracting but not the repelling fixed point of g , so $g(E) \subset E$. Using the fact that $\lambda > 40$, one can check that $g(10) \in E$ and hence $g(A) \subset E$ and $g(B) \subset E$. Next observe that, for $n = 0, 1, \dots, m$, $gf^{-1}(I_n) = f^{-1}k^{-1}(I_n)$. If $n \neq 0$, then $k^{-1}(I_n) = I_{n-1}$; hence $gf^{-1}(I_n) = f^{-1}(I_{n-1})$. For $n = 0$, we have $k^{-1}(I_0) \subset [0, 2/\lambda^{m+1}]$, so $gf^{-1}(I_0) \subset f^{-1}([0, 2/\lambda^{m+1}]) = C$. Hence $g(X) \subset X$.

Let us now prove that $h(X) \subset X$. Observe that X is contained in $\overline{\mathbb{R}} \setminus (16/5, 10)$. Using the fact that $\lambda > 40$, one can check that $h(10) \in f^{-1}(I_m)$ and hence $h(X) \subset f^{-1}(I_m)$.

Finally, we prove that $k_m(X) \subset X$. For $n = 0, 1, \dots, m-1$, we have

$$k_m(J_n) = \left[k_m^{n+1} \left(\frac{1}{\lambda^m} \right), k_m \left(\frac{2}{\lambda^{m-n}} \right) \right] = \left[k_m^{n+1} \left(\frac{1}{\lambda^m} \right), \frac{2}{\lambda^{m-(n+1)}} + \frac{2-2\lambda}{\lambda^{m+1}} \right] \subset J_{n+1}.$$

Also, $k_m(A) \subset A$, and, since

$$J_m \cup C \cup D \subset [1/2, +\infty],$$

it maps this union of intervals inside A . Hence $k_m(X) \subset X$.

We have shown that f , g , h and k_m each map X inside itself. Actually, by modifying X and k_m ever so slightly we can ensure that f , g , h and k_m each map X into the *interior* of X . To do this, we increase the size of each of the constituent intervals of X suitably, and shift the repelling fixed point of k_m to the right slightly, outside A . This can be done in such a way that f , g , h and k_m do indeed each map X into the interior of X . Hence $(f, g, h, k_m) \in \mathcal{H}$, completing our construction. \square

Motivated by this example, we briefly study discrete semigroups. The next lemma is a more general version of a well-known result that holds for Fuchsian groups (see, for example, [28, Corollary 2.2.7]). The arguments in the proof of Lemma 6.6 are borrowed from [6, Theorem 5.3.2], and we include them here in order to emphasise the fact that [6, Theorem 5.3.2] holds not only for subgroups of $\mathrm{PSL}(2, \mathbb{R})$, but for any subset of $\mathrm{PSL}(2, \mathbb{R})$.

Before we present our lemma let us recall an equivalent definition of discreteness from [6]. Let $f \in \mathrm{PSL}(2, \mathbb{R})$, with $f(z) = (az+b)/(cz+d)$. We define $\|f\| = (a^2 + b^2 + c^2 + d^2)^{1/2}$, which is a norm on $\mathrm{PSL}(2, \mathbb{R})$ (for more details, see [6, Section 2.2]). It is easy to check

that a subset A of $\mathrm{PSL}(2, \mathbb{R})$ is discrete if and only if the set

$$\{f \in A: \|f\| \leq k\},$$

is finite for all positive integers k (see [6, page 15]).

LEMMA 6.6 ([6]). *A subset A of $\mathrm{PSL}(2, \mathbb{R})$ is discrete if and only if the orbit of i under A is a discrete subset of \mathbb{H} .*

PROOF. Let $A(i)$ denote the orbit of the point i under A . Suppose first that A is discrete. Then A is a countable subset of $\mathrm{PSL}(2, \mathbb{R})$ and so we can write $A = \{f_1, f_2, \dots\}$, for some sequence (f_n) in $\mathrm{PSL}(2, \mathbb{R})$. From the characterisation of discreteness in terms of the norm on $\mathrm{PSL}(2, \mathbb{R})$ we obtain that $\|f_n\| \rightarrow \infty$, as $n \rightarrow \infty$. We now refer to [6, Theorem 4.2.1], which states that for any $f \in \mathrm{PSL}(2, \mathbb{R})$ we have

$$\|f\|^2 = 2 \cosh \rho(i, f(i)).$$

Hence, $\rho(i, f_n(i)) \rightarrow \infty$, as $n \rightarrow \infty$. This implies that the set $\{f \in A: \rho(i, f(i)) \leq k\}$ is finite for all positive integers k , which yields the desired result.

For the converse direction we prove the contrapositive: assuming that A is not discrete, we will prove that $A(i)$ is not a discrete subset of \mathbb{H} . To that end, suppose that A is not discrete. Then there exists a sequence (h_n) in A such that h_n converges to some transformation $h \in \mathrm{PSL}(2, \mathbb{R})$, and $h_n \neq h$ for all n . This implies that the sequence $(h_n(i))$ accumulates at $h(i) \in \mathbb{H}$, which concludes our proof. \square

We say that a sequence of points (z_n) in \mathbb{H} converges *conically* to $x \in \overline{\mathbb{R}}$, if $z_n \rightarrow x$, as $n \rightarrow \infty$, and there exists a hyperbolic geodesic γ landing at x and a constant $c > 0$, such that $\rho(z_n, \gamma) < c$, for all $n \in \mathbb{N}$. It is easy to see that the geodesic γ is not unique, and for two sequences that converge conically to the same boundary point we have the following.

LEMMA 6.7. *Suppose that (z_n) and (w_n) are two sequences in \mathbb{H} that converge conically to $x \in \overline{\mathbb{R}}$. Then there exists a hyperbolic geodesic γ landing at x , and a constant $c > 0$ such that*

$$\rho(z_n, \gamma) < c \quad \text{and} \quad \rho(w_n, \gamma) < c, \quad \text{for all } n \text{ large enough.}$$

PROOF. Since (z_n) and (w_n) converge conically to x , there exist geodesics γ_z and γ_w landing at x , and constants $c_z, c_w > 0$ such that

$$\rho(z_n, \gamma_z) < c_z \quad \text{and} \quad \rho(w_n, \gamma_w) < c_w, \quad \text{for all } n \in \mathbb{N}.$$

Observe that it suffices to prove that there exists a constant $c \geq \max\{c_z, c_w\}$ such that

$$\rho(w_n, \gamma_z) < c, \quad \text{for all } n \text{ large enough.}$$

Conjugating by a Möbius transformation, we may assume that γ_z is the half-line in \mathbb{H} joining 0 and infinity, and γ_w is the half-line joining 1 and infinity. Let $1 + k_n i$ be the hyperbolic orthogonal projection of w_n onto γ_w , for all n . Because (w_n) converges to the point infinity, we may assume that $k_n > 1$, for all n large enough. We then have

$$\rho(w_n, \gamma_z) \leq \rho(w_n, k_n i) \leq \rho(w_n, 1 + k_n i) + \rho(1 + k_n i, k_n i) < c_w + \rho(1 + k_n i, k_n i).$$

Hence, it suffices to prove that $\rho(1 + k_n i, k_n i)$ is bounded above, for all n large enough. Using the formulas of the hyperbolic metric in \mathbb{H} , we obtain

$$\sinh \frac{1}{2} \rho(1 + k_n i, k_n i) = \frac{|1 + k_n i - k_n i|}{2\sqrt{k_n k_n}} = \frac{1}{2k_n} \leq \frac{1}{2},$$

for all n large enough, which implies that $\rho(1 + k_n i, k_n i)$ is bounded above, as required. \square

We say that a semigroup S has *conical* limit sets, if for every point $x \in \Lambda^+(\mathcal{F})$ there exists a right-composition sequence (F_n) in $\langle \mathcal{F} \rangle$ such that $(F_n(i))$ converges conically to x , as $n \rightarrow \infty$, and similarly for every $y \in \Lambda^-(\mathcal{F})$ there exists a right-composition sequence (G_n) in $\langle \mathcal{F} \rangle^{-1}$ such that $(G_n(i))$ converges conically to y , as $n \rightarrow \infty$. Let us note that the notion of a conical limit set appears with different context in the theory of Fuchsian groups (see, for example, [39, Section 2.4]).

The final result of this chapter is the following theorem, where for a transformation h and a semigroup S , we let $h^{-1}S^{-1}h$ denote the semigroup $\{h^{-1}f^{-1}h : f \in S\}$.

THEOREM 6.8. *Suppose that $S = \langle \mathcal{F} \rangle$ is a discrete and inverse-free semigroup with conical limit sets. Then $\Lambda^+(\mathcal{F}) \cap \Lambda^-(\mathcal{F}) \neq \emptyset$ if and only if $S \cap h^{-1}S^{-1}h \neq \emptyset$, for some $h \in S$.*

PROOF. If $S \cap hS^{-1}h^{-1} \neq \emptyset$ for some $h \in S$, then there exist transformations f and g in S , such that $g = hf^{-1}h^{-1}$. Hence, $h(\alpha_f) = \beta_g$ and so β_g lies in the intersection $\Lambda^-(\mathcal{F}) \cap \Lambda^+(\mathcal{F})$.

Suppose, now, that there exists a point $x \in \Lambda^+(\mathcal{F}) \cap \Lambda^-(\mathcal{F})$. Because S has conical limit sets, we can find sequences (f_n) in S and (g_n) in S^{-1} , such that $F_n(i) = f_1 f_2 \dots f_n(i)$ and $G_n(i) = g_1 g_2 \dots g_n(i)$ converge conically to x , as $n \rightarrow \infty$. Because S is finitely-generated, by relabelling the sequences (F_n) and (G_n) , we can assume that each f_i is an ordinate of \mathcal{F} and each g_i is the inverse of an ordinate of \mathcal{F} . We are going to show that there exist subsequences (F_{n_k}) and (G_{n_k}) of (F_n) and (G_n) , respectively, such that $\rho(F_{n_k}(i), G_{n_k}(i))$ is bounded above, for all $k \in \mathbb{N}$.

Since $(F_n(i))$ and $(G_n(i))$ converge conically to the point x , Lemma 6.7 implies that there exists a constant $c > 0$ and a hyperbolic geodesic γ that lands at x , such that $\rho(F_n(i), \gamma)$ and $\rho(G_n(i), \gamma)$ are bounded above by c , for all n large enough. Let $P(F_n)$ and $P(G_n)$ denote the hyperbolic orthogonal projections of $F_n(i)$ and $G_n(i)$ onto γ , respectively. We claim that there exist subsequences (F_{n_k}) and (G_{n_k}) of (F_n) and (G_n) , respectively, such that $\rho(P(F_{n_k}), P(G_{n_k}))$ is bounded above for all $k \in \mathbb{N}$. To that end, observe that for all $n \in \mathbb{N}$ we have

$$\begin{aligned} \rho(P(F_{n+1}), P(F_n)) &\leq \rho(P(F_{n+1}), F_{n+1}(i)) + \rho(F_{n+1}(i), F_n(i)) + \rho(F_n(i), P(F_n)) \\ &< c + \rho(F_{n+1}(i), F_n(i)) + c \leq 2c + \rho(f_{n+1}(i), i) \\ &\leq 2c + \max\{\rho(f(i), i)\}, \end{aligned}$$

where the maximum is taken over all ordinates of \mathcal{F} . Let $d = 2c + \max\{\rho(f(i), i)\}$ and define the open hyperbolic discs $D(G_n(i), 2d)$, centred at $G_n(i)$ and of radius $2d$, for all n . Since $\rho(P(F_{n+1}), P(F_n))$ is bounded above by d , the disc $D(G_k(i), 2d)$ contains the point $F_{n_k}(i)$, for some positive integer n_k , and all $k \in \mathbb{N}$. Hence, there exists a subsequence (F_{n_k}) of (F_n) such that

$$\rho(P(F_{n_k}), P(G_k)) \leq 2d,$$

for all positive integers k , which proves our claim.

Now, using the triangle inequality we have that

$$\rho(F_{n_k}(i), G_k(i)) \leq \rho(F_{n_k}(i), P(F_{n_k})) + \rho(P(F_{n_k}), P(G_k)) + \rho(P(G_k), G_k(i)),$$

for all $k \in \mathbb{N}$. So

$$(6.3.1) \quad \rho(F_{n_k}(i), G_k(i)) \leq c + 4d + c,$$

for all positive integers k , as required. Because the transformations G_k preserve hyperbolic distances, (6.3.1) implies the sequence $(G_k^{-1}F_{n_k}(i))$ lies in a compact subset of \mathbb{H} . But the semigroup S is discrete, and so Lemma 6.6 yields that $G_k^{-1}F_{n_k} = h$, for some $h \in S$ and infinitely many k . Therefore, because (F_n) and (G_n) are right-composition sequences, we can find transformations f in S and g in S^{-1} so that $g^{-1}G_k^{-1}F_{n_k}f = h$, and $G_k^{-1}F_{n_k} = h$. In conclusion, we can find transformations f, h in S and g in S^{-1} such that $g^{-1}hf = h$, which yields the desired result. \square

Theorem 6.8 is reminiscent of Theorem 4.11 in the sense that it relates the size of the intersection of the limit sets of a semigroup S , to the number of conjugates of inverse elements in S . An immediate corollary of Theorem 6.8 is the following result about the cardinality of the intersection of the forward and the backward limit sets.

COROLLARY 6.9. *If $\langle \mathcal{F} \rangle$ is a discrete, inverse-free and free semigroup with conical limit sets, then the intersection $\Lambda^+(\mathcal{F}) \cap \Lambda^-(\mathcal{F})$ is at most countable.*

PROOF. From Theorem 6.8 we can find a correspondence between the sets $\Lambda^+(\mathcal{F}) \cap \Lambda^-(\mathcal{F})$ and $\langle \mathcal{F} \rangle^3$, where each $x \in \Lambda^+(\mathcal{F}) \cap \Lambda^-(\mathcal{F})$ corresponds to a triple (f, g, h) with $f = hg^{-1}h^{-1}$. Since the semigroup is free, examining the arguments in the proof of Theorem 6.8 shows that the correspondence is one-to-one. \square

It might be possible that the assumptions of Theorem 6.8 could be replaced by the assumption that $\langle \mathcal{F} \rangle$ is an inverse-free semigroup that lies in a Schottky group. This would show that the conclusion of Theorem 6.8 holds for semigroups similar to the one described in Example 6.5. Also, since Schottky groups are free groups, Corollary 6.9 would be applicable and imply that the intersection of the limit sets of a semigroup that lies in a Schottky group is “small”. This endeavour is work in progress.

CHAPTER 7

Open questions and future work

There are several open questions and problems that arise from the material presented in this thesis, which we will briefly outline in this chapter.

7.1. Right-compositions of holomorphic functions

The first problem relates to the material of Chapter 3. Suppose that (g_n) is a sequence in $\text{Hol}(\mathbb{H})$ that converges to a function $g \in \text{Hol}(\mathbb{H})$. Also, assume that the Denjoy–Wolff point of g lies in the unit circle. Example 3.9 and Theorem 3.10 indicate that there is little hope of obtaining a stability result for the right-composition sequence $G_n = g_1 g_2 \cdots g_n$. It also suggests that we ought to shift our perspective when considering right-composition sequences, in the following sense.

In Example 3.9 and Theorem 3.10, the sequence $(G_n(i))$ certainly diverges in the closure of the domain \mathbb{H} , but it converges in the Carathéodory compactification of the domain \mathbb{H} , to the prime end $[-1, 1]$. In general, for a right-composition sequence $G_n = g_1 g_2 \cdots g_n$ acting on \mathbb{H} , it is likely to be more rewarding to consider convergence of (G_n) not with respect to \mathbb{H} , but with respect to the set $\bigcap G_n(\mathbb{H})$ (or perhaps its interior), which in many cases will be a simply connected domain.

7.2. Uniform hyperbolicity

We now focus on the material presented in Chapters 5 and 6. The most important question that arises from the introduction of the semidiscrete locus is the following:

QUESTION 7.1. Is $\mathcal{S} \setminus \overline{H_P}$ equal to the closure of $\mathcal{H} \setminus H_P$ in $\text{PSL}(2, \mathbb{R})^N$?

A positive answer to Question 7.1 would imply that Theorem 5.27 provides a complete characterisation of the complement of $\overline{\mathcal{H}}$. The inclusion $\overline{\mathcal{H} \setminus H_P} \subseteq \mathcal{S} \setminus \overline{H_P}$ is obvious from the fact that $\mathcal{H} \setminus H_P$ is contained in $\mathcal{S} \setminus \overline{H_P}$ and $\mathcal{S} \setminus \overline{H_P}$ is closed. Hence, the difficulty

in Question 7.1 lies in evaluating whether every N -tuple in \mathcal{S} that does not generate a Schottky semigroup of rank one, lies in the closure of \mathcal{H} .

A first step in answering Question 7.1 would be to focus on N -tuples that generate Schottky semigroups. Due to Lemma 5.16, we only need to consider N -tuples that generate Schottky semigroups of rank strictly greater than one, since all others are contained in the closure of the principal component. So we ask the following:

QUESTION 7.2. Let \mathcal{F} be an N -tuple that generates a Schottky semigroup of rank greater than one. Does \mathcal{F} lie in the closure of $\mathcal{H} \setminus H_P$?

Recall Theorem 6.3, which states that all N -tuples in the closure of a non-principal component generate Schottky semigroups of rank greater than one. This, together with Lemma 6.2 and the fact that the intersection of the limit sets of a Schottky semigroup is finite, seem to indicate that the answer to Question 7.2 is positive. In order to provide this answer, one would have to perturb the ordinates of the generating N -tuple in such a way that the intersection points of the limit sets vanish.

Theorem 6.8 and Corollary 6.9 indicate that N -tuples that generate semigroups that lie in a Schottky group exhibit behaviour similar to that described in Lemma 6.2, but with a countable intersection of the limit sets. So it is likely that if one were to answer Question 7.2 positively, one would also be able to prove that every N -tuple that generates a semigroup which is contained in Schottky group lies in the closure of \mathcal{H} .

What is important about Question 7.2 is that a positive answer would imply that points of \mathcal{S} that lie in the interior of $\overline{\mathcal{E}_I}$, if those indeed exist, have to generate semigroups with “wild” limit sets. However, a positive answer to Question 7.2 still does not provide an answer to Question 7.1. Instead, we examine the following:

QUESTION 7.3. Suppose that \mathcal{F} is an N -tuple in \mathcal{S} . Can we find a sequence (\mathcal{F}_n) in \mathcal{S} such that $\mathcal{F}_n \rightarrow \mathcal{F}$, as $n \rightarrow \infty$, and $\langle \mathcal{F}_n \rangle$ is a Schottky semigroup?

Question 7.3 is equivalent to examining whether every finitely generated, semidiscrete and inverse-free semigroup is the algebraic limit of Schottky semigroups (see [34, Chapter 4] for the definition of algebraic convergence). This can be thought of as an analogue of the density theorem for Kleinian groups that was conjectured by Bers, Sullivan and

Thurston in the 70's, and proved recently by Ohshika [40]. The density theorem states that any discrete subgroup of $\mathrm{PSL}(2, \mathbb{C})$ (the group of complex Möbius transformations) is an algebraic limit of simpler discrete groups whose fundamental polygon has finitely many sides, called geometrically finite groups.

In our context, a positive answer to Questions 7.2 and 7.3 would provide a positive answer to Question 7.1. Also, due to the fact that $\mathcal{S} \setminus \overline{H_P}$ is closed, a counterexample to Question 7.3 would have to be a semigroup with limit sets that intersect in a convoluted way, and a generating N -tuple that is isolated from the hyperbolic components of \mathcal{H} . This would most likely mean that the N -tuple has to lie in the interior of $\overline{\mathcal{E}_I}$, thus also providing a counterexample to Question 7.1.

Bibliography

- [1] M. Abate, *Iteration theory of holomorphic maps on taut manifolds*, Research and Lecture Notes in Mathematics. Complex Analysis and Geometry, Mediterranean Press, Rende, 1989.
- [2] A. Avila, J. Bochi, and J.-C. Yoccoz, *Uniformly hyperbolic finite-valued $\mathrm{SL}(2, \mathbb{R})$ -cocycles*, Comment. Math. Helv. **85** (2010), no. 4, 813–884.
- [3] I. N. Baker and Ch. Pommerenke, *On the iteration of analytic functions in a halfplane. II*, J. London Math. Soc. (2) **20** (1979), no. 2, 255–258.
- [4] I. N. Baker and P. J. Rippon, *Towers of exponents and other composite maps*, Complex Variables Theory Appl. **12** (1989), no. 1-4, 181–200.
- [5] B. Bárány, A. Käenmäki, and H. Koivusalo, *Dimension of self-affine sets for fixed translation vectors*, J. Lond. Math. Soc. (2) **98** (2018), no. 1, 223–252.
- [6] A. F. Beardon, *The geometry of discrete groups*, Graduate Texts in Mathematics, vol. 91, Springer-Verlag, New York, 1995. Corrected reprint of the 1983 original.
- [7] ———, *Continued fractions, discrete groups and complex dynamics*, Comput. Methods Funct. Theory **1** (2001), 535–594.
- [8] ———, *Semi-groups of analytic maps*, Comput. Methods Funct. Theory **1** (2001), 249–258.
- [9] A. F. Beardon, T. K. Carne, D. Minda, and T. W. Ng, *Random iteration of analytic maps*, Ergodic Theory Dynam. Systems **24** (2004), no. 3, 659–675.
- [10] A. F. Beardon and D. Minda, *The hyperbolic metric and geometric function theory*, Quasiconformal mappings and their applications, Narosa, New Delhi, 2007, pp. 9–56.
- [11] A. F. Beardon and I. Short, *Norms of Möbius maps*, Bull. Lond. Math. Soc. **42** (2010), no. 3, 499–505.
- [12] J. Bochi and N. Gourmelon, *Some characterizations of domination*, Math. Z. **263** (2009), no. 1, 221–231.
- [13] J. Bochi and M. Rams, *The entropy of Lyapunov-optimizing measures of some matrix cocycles*, J. Mod. Dyn. **10** (2016), 255–286.
- [14] P. S. Bourdon, V. Matache, and J. H. Shapiro, *On convergence to the Denjoy-Wolff point*, Illinois J. Math. **49** (2005), no. 2, 405–430.
- [15] D. M. Burns and S. G. Krantz, *Rigidity of holomorphic mappings and a new Schwarz lemma at the boundary*, J. Amer. Math. Soc. **7** (1994), no. 3, 661–676.

BIBLIOGRAPHY

- [16] L. Carleson and T. W. Gamelin, *Complex dynamics*, Universitext: Tracts in Mathematics, Springer-Verlag, New York, 1993.
- [17] A. Christodoulou and I. Short, *A hyperbolic-distance inequality for holomorphic maps*, Ann. Acad. Sci. Fenn. Math. **44** (2019), no. 1, 293–300.
- [18] A. Christodoulou and I. Short, *Stability of the Denjoy–Wolff theorem*, available at <https://arxiv.org/abs/1907.09366>.
- [19] A. Christodoulou, *Universal constraints on semigroups of hyperbolic isometries*, available at <https://arxiv.org/abs/1907.00347>.
- [20] M. D. Contreras, S. Díaz-Madrigal, and C. Pommerenke, *Boundary behavior of the iterates of a self-map of the unit disk*, J. Math. Anal. Appl. **396** (2012), no. 1, 93–97.
- [21] O. Forster, *Lectures on Riemann surfaces*, Graduate Texts in Mathematics, vol. 81, Springer-Verlag, New York, 1991. Translated from the 1977 German original by Bruce Gilligan; Reprint of the 1981 English translation.
- [22] D. Fried, S. M. Marotta, and R. Stankewitz, *Complex dynamics of Möbius semigroups*, Ergodic Theory Dynam. Systems **32** (2012), no. 6, 1889–1929.
- [23] J. Gill, *Compositions of analytic functions of the form $F_n(z) = F_{n-1}(f_n(z))$, $f_n(z) \rightarrow f(z)$* , J. Comput. Appl. Math. **23** (1988), no. 2, 179–184.
- [24] ———, *Complex dynamics of the limit periodic system $F_n(z) = F_{n-1}(f_n(z))$, $f_n \rightarrow f$* , J. Comput. Appl. Math. **32** (1990), no. 1-2, 89–96. Extrapolation and rational approximation (Luminy, 1989).
- [25] A. Hinkkanen and G. J. Martin, *The dynamics of semigroups of rational functions. I*, Proc. London Math. Soc. (3) **73** (1996), no. 2, 358–384.
- [26] M. Jacques and I. Short, *Dynamics of hyperbolic isometries*, available at <https://arxiv.org/abs/1609.00576v4>.
- [27] A. Karlsson, *Non-expanding maps and Busemann functions*, Ergodic Theory Dynam. Systems **21** (2001), no. 5, 1447–1457.
- [28] S. Katok, *Fuchsian groups*, Chicago Lectures in Mathematics, University of Chicago Press, Chicago, IL, 1992.
- [29] L. Keen and N. Lakic, *Random holomorphic iterations and degenerate subdomains of the unit disk*, Proc. Amer. Math. Soc. **134** (2006), no. 2, 371–378.
- [30] ———, *Hyperbolic geometry from a local viewpoint*, London Mathematical Society Student Texts, vol. 68, Cambridge University Press, Cambridge, 2007.
- [31] J. L. Kelley, *General topology*, D. Van Nostrand Company, Inc., Toronto-New York-London, 1955.
- [32] A. Kuznetsov, *Semi-groups of analytic functions that contain the identity map*, Comput. Methods Funct. Theory **7** (2007), no. 1, 239–247.
- [33] L. Lorentzen, *Compositions of contractions*, J. Comput. Appl. Math. **32** (1990), no. 1-2, 169–178. Extrapolation and rational approximation (Luminy, 1989).

BIBLIOGRAPHY

- [34] A. Marden, *Outer circles*, Cambridge University Press, Cambridge, 2007. An introduction to hyperbolic 3-manifolds.
- [35] B. Maskit, *A characterization of Schottky groups*, J. Analyse Math. **19** (1967), 227–230.
- [36] ———, *Kleinian groups*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 287, Springer-Verlag, Berlin, 1988.
- [37] J. Milnor, *Dynamics in one complex variable*, 3rd ed., Annals of Mathematics Studies, vol. 160, Princeton University Press, Princeton, NJ, 2006.
- [38] R. Narasimhan, *Several complex variables*, The University of Chicago Press, Chicago, Ill.-London, 1971. Chicago Lectures in Mathematics.
- [39] P. J. Nicholls, *The ergodic theory of discrete groups*, London Mathematical Society Lecture Note Series, vol. 143, Cambridge University Press, Cambridge, 1989.
- [40] K. Ohshika, *Realising end invariants by limits of minimally parabolic, geometrically finite groups*, Geom. Topol. **15** (2011), no. 2, 827–890.
- [41] Ch. Pommerenke, *On asymptotic iteration of analytic functions in the disk*, Analysis **1** (1981), no. 1, 45–61.
- [42] ———, *Boundary behaviour of conformal maps*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 299, Springer-Verlag, Berlin, 1992.
- [43] J. L. Schiff, *Normal families*, Universitext, Springer-Verlag, New York, 1993.
- [44] J. H. Shapiro, *Composition operators and classical function theory*, Universitext: Tracts in Mathematics, Springer-Verlag, New York, 1993.
- [45] K. Tavakoli, *Boundary points as limit functions of iterated holomorphic function systems*, Proc. Amer. Math. Soc. **137** (2009), no. 6, 1971–1976.
- [46] J.-C. Yoccoz, *Some questions and remarks about $SL(2, \mathbf{R})$ cocycles*, Modern dynamical systems and applications, Cambridge Univ. Press, Cambridge, 2004, pp. 447–458.
- [47] A. Zimmer, *Two boundary rigidity results for holomorphic maps*, available at <https://arxiv.org/abs/1810.05669v2>.